PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
A01H 5/00, 5/10, A01N 37/18, C07K 4/10, 7/08, 14/415, C12N 15/29

A1

(11) International Publication Number: **WO 98/27805**

(43) International Publication Date: 2 July 1998 (02.07.98)

(21) International Application Number: PCT/AU97/00874

(22) International Filing Date: 22 December 1997 (22.12.97)

(30) Priority Data:
PO 4275 20 December 1996 (20.12.96) AU

(71) Applicant *(for all designated States except US)*: COOPERATIVE RESEARCH CENTRE FOR TROPICAL PLANT PATHOLOGY [AU/AU]; The University of Queensland, John Hines Building, 5th Level, St. Lucia, QLD 4067 (AU).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: MANNERS, John, Michael [AU/AU]; 28 Warmington Street, Paddington, QLD 4064 (AU). MARCUS, John, Paul [US/AU]; 1–10 Balderstone Street, Corinda, QLD 4075 (AU). GOULTER, Kenneth, Clifford [AU/AU]; 26 Emblem Street, Jamboree Heights, QLD 4074 (AU). GREEN, Jodie, Lyn [AU/AU]; 4 Reef Close, Jamboree Heights, QLD 4074 (AU). BOWER, Neil, Ivan [AU/AU]; 22 Fallbrook Street, Kenmore, QLD 4069 (AU).

(74) Agent: CULLEN & CO.; Level 12, 240 Queen Street, Brisbane, QLD 4000 (AU).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: ANTIMICROBIAL PROTEINS

(57) Abstract

A new family of antimicrobial proteins is described. Prototype proteins can be isolated from *Macadamia integrifolia* as well as other plant species. DNA encoding the protein is also described as well as DNA constructs which can be used to express the antimicrobial protein or to introduce the antimicrobial protein into a plant. Compositions comprising the antimicrobial proteins or the antimicrobial protein *per se* can be administered to plants or mammilian animals to combat microbial infestation.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

ANTIMICROBIAL PROTEINS

TECHNICAL FIELD

This invention relates to isolated proteins which exert inhibitory activity on the growth of fungi and bacteria, which fungi and bacteria include some microbial pathogens of plants and animals. The invention also relates to recombinant genes which include sequences encoding the proteins, the expression products of which recombinant genes can contribute to plant cells or cells of other organism's defence against invasion by microbial pathogens. The invention further relates to the use of the proteins and/or genes encoding the proteins for the control of microbes in human and veterinary clinical conditions.

BACKGROUND ART

Microbial diseases of plants are a significant problem to the agricultural and horticultural industries. Plant diseases in general cause millions of tonnes of crop losses annually with fungal and bacterial diseases responsible for significant portions of these losses. One possible way of combating fungal and bacterial diseases is to provide transgenic plants capable of expressing a protein or proteins which in some way increase the resistance of the plant to pathogen attack. A simple strategy is to first identify a protein with antimicrobial activity *in vitro*, to clone or synthesise the DNA sequence encoding the protein, to make a chimaeric gene construct for efficient expression of the protein in plants, to transfer this gene to transgenic plants and to assess the effect of the introduced gene on resistance to microbial pathogens by comparison with control plants.

The first and most important step in the strategy for disease control described above is to identify, characterise and describe a protein with strong antimicrobial activity. In recent years, many different plant proteins with antimicrobial and/or antifungal activity have been identified and described. These proteins have been categorised into several classes according to either their presumed mode of action and/or their amino acid sequence homologies. These classes include the following: chitinases (Roberts, W.K. *et al.* [1986] *Biochim. Biophys. Acta* 880:161-170); β-1,3-glucanases (Manners, J.D. *et al.* [1973] *Phytochemistry* 12:547-553); thionins (Bolmann, H. *et al.* [1988] *EMBO J.* 7:1559-1565 and Fernadez de Caleya, R. *et al.* [1972] *Appl. Microbiol.* 23:998-1000); permatins (Roberts, W. K. *et al.* [1990] *J. Gen. Microbiol.* 136:1771-1778 and Vigers, A.J. *et al.* [1991] *Mol. Plant-Microbe Interact.* 4:315-323); ribosome-inactivating proteins (Roberts, W. K. *et al.* [1986] *Biochim. Biophys. Acta* 880:161-170 and Leah, R. *et al.* [1991] *J. Biol. Chem.* 266:1564-1573); plant defensins (Terras, F. R. G. *et al.* [1995] *The Plant Cell* 7:573-588); chitin binding proteins (De Bolle, M.F.C. *et al.* [1992] *Plant Mol. Biol.* 22:1187-1190 and Van Parijs, J. *et al.* [1991] *Planta* 183:258-264); thaumatin-like, or osmotin-like proteins (Woloshuk, C.P. *et al.* [1991] *The Plant Cell* 3:619-628 and Hejgaard, J. [1991] *FEBS Letts.* 291:127-131); PR1-type

proteins (Niderman, T. *et al.* [1995] *Plant Physiol.* 108:17-27.) and the non-specific lipid transfer proteins (Terras, F.R.G. *et al.* [1992] *Plant Physiol.* 100:1055-1058 and Molina, A. *et al.* [1993] *FEBS Letts.* 3166:119-122). Another class of antimicrobial proteins from plants is the knottin or knottin-like antimicrobial proteins (Cammue, B.P.A. *et al.* [1992] *J. Biol. Chem.* 67:2228-2233; Broekaert W.F. *et al.* (1997) *Crit. Rev. in Plant Sci.* 16(3):297-323). A class of antimicrobial proteins termed 4-cysteine proteins has also been reported in the literature which class includes Maize Basic Protein (MBP-1) (Duvick, J.P. *et al.* [1992] J. Biol. Chem. 267:18114-18120). A novel antimicrobial protein which does not fit into any previously described class of antimicrobial proteins has also been isolated from the seeds of *Macadamia integrifolia* termed MiAMP1 (Marcus, J.P. *et al.* [1997] *Eur. J. Biochem.* 244:743-749). In addition, plants are not the sole source of antimicrobial proteins and there are many reports of the isolation of antimicrobial proteins from animal and microbial cells (reviewed in Gabay, J.E. [1994] *Science* 264:373-374 and in "Antimicrobial peptides" [1994] *CIBA Foundation Symposium 186*, John Wiley and Sons Publ., Chichester, UK).

There is evidence that the ectopic expression of genes encoding proteins that have *in vitro* antimicrobial activity in transgenic plants can result in increased resistance to microbial pathogens. Examples of this engineered resistance include transgenic plants expressing genes encoding: a plant chitinase, either alone (Broglie, K. *et al.* [1991] *Science* 254:1194-1197) or in combination with a β-1,3-glucanase (Van den Elzen, P.J.M. *et al.* [1993] *Phil. Trans. Roy. Soc.* 342:271-278); a plant defensin (Terras, F.R.G. *et al.* [1995] *The Plant Cell* 7:573-588); an osmotin-like protein (Liu, D. *et al.* [1994] *Proc. Natl. Acad. Sci. USA* 91:1888-1892); a PR1-class protein (Alexander, D. *et al.* [1993] *Proc. Natl. Acad. Sci. USA* 90:7327-7331) and a ribosome-inactivating protein (Logemann, J. *et al.* [1992] *Bio/Technology* 10:305-308).

Although the potential use of antimicrobial proteins for engineering disease resistance in transgenic plants has been described extensively, there are other applications which are worthy of mention. Firstly, highly potent antimicrobial proteins can be used for the control of plant disease by direct application (De Bolle, M.F.C. *et al.* [1993] in *Mechanisms of Plant Defence Responses*, B. Fritig and M. Legrand eds., Kluwer Acad. Publ., Dordrecht, NL, pp. 433-436). In addition, antimicrobial peptides have potential therapeutic applications in human and veterinary medicine. Although this has not been described for peptides of plant origin it is being actively explored with peptides from animals and has reached clinical trials (Jacob, L. and Zasloff, M. [1994] in "Antimicrobial Peptides", *CIBA Foundation Symposium 186*, John Wiley and Sons Publ., Chichester, UK, pp. 197-223).

Antimicrobial proteins exhibit a variety of three-dimensional structures which will determine in large part the activity which they manifest. Many of the global structures exhibited by these

proteins have been determined (Broekaert W.F. *et al.* (1997) *Crit. Rev. in Plant Sci.* 16(3):297-323). A large factor in determining the stability of these proteins is the presence of disulfide bridges between various cysteines located in α-helical and β-sheet regions. Many peptides with toxic activity such as conotoxin are well known to be stabilized by disulfide bridges (see for example Hill, J.M. *et al.* (1996) *Biochemistry* 35(27): 8824-8835). In the case of the conotoxin referenced above, a compact structure is formed consisting of a helix, a small -hairpin, a cis-hydroxyproline, and several turns. The molecule is stabilized by three disulfide bonds, two of which connect the α-helix and the β-sheet, forming a solid structural core. Interestingly, eight arginine and lysine side chains in this molecule project into the solvent in a radial orientation relative to the core of the molecule. These cationic side chains form potential sites of interaction with anionic sites on pathogen membranes (Hill, J.M. *et al. supra*).

The invention described herein constitutes previously undiscovered and thus novel proteins with antimicrobial activity. These proteins can be isolated from *Macadamia integrifolia* (Mi) seeds or from cotton or cocoa seeds. In addition, protein fragments which are antifungal can be derived from larger seed storage proteins containing regions of substantial similarity to the antimicrobial proteins from macadamia described here. Examples of seed storage proteins which contain regions similar to the proteins which have been purified can be seen in Figure 4. *Macadamia integrifolia* belongs to the family Proteaceae. *M. integrifolia*, also known as Bauple Nut or Queensland Nut, is considered by some to be the world's best edible nut. Cotton (*Gossypium hirsutum*) belongs to the family Malvaceae and is cultivated extensively for its fiber. Cocoa (*Threobroma cacao*) belongs to the family Sterculiaceae and is used around the world for a wide variety of cocoa products.

The fact that both the macadamia and cocoa antimicrobial proteins are found in edible portions of these plants makes these peptides attractive for use in genetic engineering for disease resistance since trangenic plants expressing these proteins are unlikely to show added toxicity. Proteins may also be safe for human and veterinary use.

SUMMARY OF THE INVENTION

According to a first embodiment of the invention, there is provided a protein fragment having antimicrobial activity, wherein said protein fragment is selected from:

(i) a polypeptide having an amino acid sequence selected from:

residues 29 to 73 of SEQ ID NO: 1

residues 74 to 116 of SEQ ID NO: 1

residues 117 to 185 of SEQ ID NO: 1

residues 186 to 248 of SEQ ID NO: 1

residues 29 to 73 of SEQ ID NO: 3

residues 74 to 116 of SEQ ID NO: 3
residues 117 to 185 of SEQ ID NO: 3
residues 186 to 248 of SEQ ID NO: 3
residues 1 to 32 of SEQ ID NO: 5
residues 33 to 75 of SEQ ID NO: 5
residues 76 to 144 of SEQ ID NO: 5
residues 145 to 210 of SEQ ID NO: 5
residues 34 to 80 of SEQ ID NO: 7
residues 81 to 140 of SEQ ID NO: 7
residues 33 to 79 of SEQ ID NO: 8
residues 80 to 119 of SEQ ID NO: 8
residues 120 to 161 of SEQ ID NO: 8
residues 32 to 91 of SEQ ID NO: 21
residues 25 to 84 of SEQ ID NO: 22
residues 29 to 94 of SEQ ID NO: 24
residues 31 to 85 of SEQ ID NO: 25
residues 1 to 23 of SEQ ID NO: 26
residues 1 to 17 of SEQ ID NO: 27
residues 1 to 28 of SEQ ID NO: 28;

(ii) a homologue of (i);

(iii) a polypeptide containing a relative cysteine spacing of C-2X-C-3X-C-(10-12)X-C-3X-C-3X-C wherein X is any amino acid residue, and C is cysteine;

(iv) a polypeptide containing a relative cysteine and tyrosine/phenylalanine spacing of Z-2X-C-3X-C-(10-12)X-C-3X-C-3X-Z wherein X is any amino acid residue, and C is cysteine, and Z is tyrosine or phenylalanine;

(v) a polypeptide containing a relative cysteine spacing of C-3X-C-(10-12)X-C-3X-C wherein X is any amino acid residue, and C is cysteine;

(vi) a polypeptide with substantially the same spacing of positively charged residues relative to the spacing of cysteine residues as (i); and

(vii) a fragment of the polypeptide of any one of (i) to (vi) which has substantially the same antimicrobial activity as (i).

According to a second embodiment of the invention, there is provided a protein containing at least one polypeptide fragment according to the first embodiment, wherein said polypeptide fragment

has a sequence selected from within a sequence comprising SEQ ID NO: 1, SEQ ID NO: 3 or SEQ ID NO: 5.

According to a third embodiment of the invention, there is provided a protein having a sequence selected from SEQ ID NO: 1, SEQ ID NO: 3 or SEQ ID NO: 5.

According to a fourth embodiment of the invention, there is provided an isolated or synthetic DNA encoding a protein according to the first embodiment

According to a fifth embodiment of the invention, there is provided a DNA construct which includes a DNA according to the fourth embodiment operatively linked to elements for the expression of said encoded protein.

According to a sixth embodiment of the invention, there is provided a transgenic plant harbouring a DNA construct according to the fifth embodiment.

According to a seventh embodiment of the invention, there is provided reproductive material of a transgenic plant according to the sixth embodiment.

According to an eighth embodiment of the invention, there is provided a composition comprising an antimicrobial protein according to the first embodiment together with an agriculturally-acceptable carrier diluent or excipient.

According to a ninth embodiment of the invention, there is provided a composition comprising an antimicrobial protein according to the first embodiment together with an pharmaceutically-acceptable carrier diluent or excipient.

According to a tenth embodiment of the invention, there is provided a method of controlling microbial infestation of a plant, the method comprising:

i) treating said plant with an antimicrobial protein according to the first embodiment or a composition according to the eighth embodiment; or

ii) introducing a DNA construct according to the fifth embodiment into said plant.

According to an eleventh embodiment of the invention, there is provided a method of controlling microbial infestation of a mammalian animal, the method comprising treating the animal with an antimicrobial protein according to the first embodiment or a composition according to the ninth embodiment.

According to a twelfth embodiment of the invention, there is provided a method of preparing an antimicrobial protein, which method comprises the steps of:

a) obtaining or designing an amino acid sequence which forms a helix-turn-helix structure;

b) replacing individual residues to achieve substantially the same distribution of positively charged residues and cysteine residues as in one or more of the amino acid sequences shown in Figure 4;

c) synthesising a protein comprising said amino acid sequence chemically or by recombinant DNA techniques in liquid culture; and

d) if necessary, forming disulphide linkages between said cysteine residues.

Other embodiments of the invention include methods for producing antimicrobial protein.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the results of cation-exchange chromatography of the basic protein fraction of a *Macadamia integrifolia* extract with the results of a bioassay for antimicrobial activity shown for fractions in the region of MiAMP2c elution.

Figure 2 shows the results of including 1 mM Ca^{2+} in a parallel bioassay of fractions from the cation-exchange separation.

Figure 3 shows a reverse-phase HPLC profile of highly inhibitory fractions containing MiAMP2c from the cation-exchange separation in Figure 1 and 2 together with % growth inhibition exhibited by the HPLC fractions.

Figure 4 shows the amino acid sequences of MiAMP2a, b, c and d and protein fragments derived from other seed storage proteins which contain regions of homology to the MiAMP2 series of antimicrobial proteins.

Figure 5 shows an example of a synthetic nucleotide sequence which can be used for the expression and secretion of MiAMP2c in transgenic plants.

Figure 6 shows the alignment of clones 1-3 from macadamia containing MiAMP2a, b, c and d subunits together with sequences from cocoa and cotton vicilin seed storage proteins which exhibit significant homology to the macadamia clones.

Figure 7 displays a series of secondary structure predictions for MiAMP2c.

Figure 8 shows a three-dimensional model of the MiAMP2c protein.

Figure 9 shows stained SDS-PAGE gels of protein fractions at various stages in the expression and purification of TcAMP1(Theobroma cacao subunit 1), MiAMP2a, MiAMP2b, MiAMP2c and MiAMP2d expressed in *E.coli* liquid culture.

Figure 10 shows the reverse-phase HPLC purification of cocoa subunit 2 (TcAMP2) after the initial purification step using Ni-NTA media.

Figure 11 shows a western blot of crude protein extracts from various plant species using rabbit antiserum raised to MiAMP2c.

Figure 12 shows a cation-exchange fractionation of the *Stenocarpus sinuatus* basic protein fraction along with the accompanying western blot which shows the presence of immunologically-related proteins in a range of fractions.

Figure 13 shows a reverse-phase HPLC separation of *Stenocarpus sinuatus* cation-exchange fractions which had previously reacted with MiAMP2c antibodies (see Figure 14). A western blot is also presented which reveals the presence of putative MiAMP2c homologues in individual HPLC fractions.

Figure 14 is a map of the binary vector pPCV91-MiAMP2c as an example of a vector that can be used to express these antimicrobial proteins in transgenic plants.

Figure 15 shows a western blot to detect MiAMP2c expressed in transgenic tobacco plants.

BEST MODE AND OTHER MODES FOR CARRYING OUT THE INVENTION

The following abbreviations are used hereafter:

EDTA	ethylenediaminetetraacetic acid
IPTG	Isopropyl-β-D-thiogalactopyranoside
MeCN	methyl cyanide (acetonitrile)
Mi	*Macadamia integrifolia*
MiAMP2	*Macadamia integrifolia* antimicrobial protein series number 2
Ni-NTA	Nickel-nitrilotriacetic acid chromatography media
ND	not determined
PCR	polymerase chain reaction
PMSF	phenylmethylsulphonyl fluoride
SDS-PAGE	sodium-dodecylsulphate polyacrylamide gel electrophoresis
TFA	trifluoroacetate

The term homologue is used herein to denote any polypeptide having substantial similarity in composition and sequence to the polypeptide used as the reference. The homologue of a reference polypeptide will contain key elements such as cysteine or other residues spaced at identical intervals such that a substantially similar three-dimensional global structure is adopted by the homologue as compared to the reference. The homologue will also exhibit substantially the same antimicrobial activity as the reference protein.

The present inventors have identified a new class of proteins with antimicrobial activity. Prototype proteins can be isolated from seeds of *Macadamia integrifolia*. The invention thus provides antimicrobial proteins *per se* and also DNA sequences encoding these antimicrobial proteins.

The invention also provides amino acid sequences of proteins which are homologous to the prototype antimicrobial proteins from *Macadamia integrifolia*. Thus, in addition to the antimicrobial proteins from Macadamia, this invention also provides amino acid sequences of homologues from other species which have hitherto been unrecognized as having antimicrobial activity.

While the first antimicrobial protein in the present series was isolated directly from *Macadamia integrifolia*, additional antimicrobial proteins were identified through cloning efforts, homology searches and subsequent antimicrobial testing of the encoded proteins after expression in and purification from liquid culture. After the first protein from this series was purified from macadamia and termed MiAMP2, clones were obtained which encoded a preproprotein containing MiAMP2. This large protein (666 amino acids), represented by several almost identical clones, contained four adjacent regions with significant similarity to the purified antimicrobial protein fragment (MiAMP2) which itself was found to lie within region three in the cloned nucleotide sequence; hence the purified antimicrobial protein is termed MiAMP2c. Other fragments contained in the 666-amino-acid clone are termed MiAMP2a, b and d as per their locations in the cloned nucleotide sequence. Several other sequences with significant homology to the MiAMP2a, b, c, and d protein fragments were then identifed in the Entrez data base. These homologous sequences were contained within larger seed storage proteins from cotton and cocoa which sequences had not been previously described as containing antimicrobial protein sequences or as exhibiting antimicrobial activity. Fragments of larger seed storage proteins containing sequences homologous to MiAMP2c were tested and are here demonstrated to exhibit antimicrobial activity. Thus, the inventors have established a process for obtaining antimicrobial protein fragments from larger seed storage proteins. In the light of these findings, it is evident that fragments of other seed storage proteins containing sequences similar to the proteins described will also exhibit antimicrobial activity.

In particular, the 47-amino-acid TcAMP1 (for *Theobroma cacao* antimicrobial protein 1) and the 60-amino-acid TcAMP2 sequences were derived from a cocoa vicilin seed storage protein gene sequence (which contains 525 amino acids) (Spencer, M.E. and Hodge R. [1992] *Planta* 186:567-576). These derived fragments were then expressed in liquid culture. Cocoa vicilin fragments thus expressed and subsequently purified (Examples 10 and 11), were shown to be antimicrobial (Example 15). This is the first report that fragments of the cocoa vicilin protein possess antimicrobial activity. Pools of sequences containing fragments homologous to the MiAMP2c apparently released from cotton vicilin seed storage protein have been shown to possess antimicrobial activity (Chung, R. P.T. *et al.* [1997] *Plant Science* 127:1-16). This finding is clearly embodied in sequences disclosed in this application.

In addition to showing that cocoa-vicilin-derived fragments exhibit antimicrobial activity, there is herein described additional proteins which exhibit antimicrobial activity. For example, there is described below proteins from *Stenocarpus sinuatus* which are of similar size to MiAMP2 subunits, react with MiAMP2c antiserum, and contain sequences homologous to MiAMP2 proteins (see Figure 4). Based on the evidence provided herein, sequences homologous to the MiAMP2c

subunit (i.e., MiAMP2a, b, d; TcAMP1; TcAMP2; and cotton fragments 1, 2 and 3—see Figure 4) constitute proteins which contain the fragment with antimicrobial activity. The antimicrobial activity of MiAMP2 fragments from macadamia, and the TcAMP1 and 2 fragments from cocoa, is exemplified below. R. P. T. Chung *et al.* (*Plant Science* 127:1-16 [1997]) have demonstrated that the cotton fragments exhibit antimicrobial activity. Other antimicrobial proteins can also be derived from seed storage proteins such as peanut allergen Ara h (Burks, A.W. *et al.* [1995] *J. Clin. Invest.* 96 (4), 1715-1721), maize globulin (Belanger, F. C. and Kriz, A. L.[1991] *Genetics* 129 (3), 863-872), barley globulin (Heck, G. R. *et al.* [1993] *Mol. Gen. Genet.* 239 (1-2), 209-218), and soybean conglycinin (Sebastiani, F. L. *et al.* [1990] *Plant Mol. Biol.* 15 (1), 197-201), all of which contain the same key elements which are present in the sequences which are here shown to exhibit antimicrobial activity.

The proteins which contain regions of sequence homologous to MiAMP2 (as in Figure 4) can be used to construct nucleotide sequences encoding 1) the active fragments of larger proteins, or 2) fusions of multiple antimicrobial fragments. This can be done using standard codon tables and cloning methods as described in laboratory manuals such as *Current Protocols in Molecular Biology* (copyright 1987-1995 edited by Ausubel F. M. *et al.* and published by John Wiley & Sons, Inc., printed in the USA). Subsequently, these can be expressed in liquid culture for purification and testing, or the sequences can be expressed in transgenic plants after placing them in appropriate expression vectors.

The antimicrobial proteins *per se* will manifest a particular three-dimensional structure which may be determined using X-ray crystallography or nuclear magnetic resonance techniques. This structure will be responsible in large part for the antimicrobial activity of the protein. The sequence of the protein can also be subjected to structure prediction algorithms to assess whether any secondary structure elements are likely to be exhibited by the protein (see Example 8 and Figure 7). Secondary structures, thus predicted, can then be used to model three-dimensional global structures. Although three-dimensional structure prediction is not feasible for most proteins, the secondary structure predictions for MiAMP2c were sufficiently simple and clear that a three-dimensional model structure has been obtained for the MiAMP2c protein. Homologues exhibiting the same cysteine spacing and other key elements will also adopt the same three-dimensional structure. Example 8 shows that the structure most likely to be adopted by MiAMP2c (and homologues) is a helix-turn-helix structure stabilised by at least two disulfide bridges connecting the two antiparallel α-helical segments (see Figure 8). Additional stabilisation can be provided by an extra disulfide bridge (e.g., as in MiAMP2b) or by a hydrophobic ring-stacking interaction between tyrosine and/or phenylalanine residues (e.g., MiAMP2a and MiAMP2c), each located on the same face of the α-

helical segments as the normally present cysteine residues which participate in the 2 disulfide linkages mentioned above. NMR signals exhibited by MiAMP2c are consistent with the three-dimensional global model produced from the secondary-structure predictions mentioned above.

It will be appreciated that one skilled in the art could take a protein with known structure, alter the sequence significantly, and yet retain the overall three-dimensional shape and antimicrobial activity of the protein. One aspect of the structure that most likely could not be altered without seriously affecting the structure (and, therefore, the activity of the protein) is the content and spacing of the cysteine residues since this would disrupt the formation of disulfide bonds which are critical to a) maintaining the overall structure of the protein and/or b) making the protein more resistant to denaturation and proteolysis (stabilizing the protein structure). In particular, it is essential that cysteine residues reside on one face of the helix in which they are contained. This can best be accomplished by maintaining a three-residue spacing between the cysteine residues within each helix, but, can also be accomplished with a two-residue interval between the cysteine residues - provided the cysteines on the other helical segment are separated by three residues (i.e., C-X-X-C-X-X-X-C-nX-C-X-X-X-C-X-X-X-C where C is cysteine, X is any amino acid, and n is the number of residues forming a turn between the two α-helical segments). Aromatic tyrosine (or phenylalanine) residues can also function to add stability to the protein structure if they are located on the same face of the helix as the cysteine side chains. This can be accomplished by providing appropriate spacing of two or three residues between the aromatic residue and the proximate cysteine residue (i.e., Z-X-X-C-X-X-X-C-nX-C-X-X-X-C-X-X-X-Z where Z is tyrosine or phenylalanine).

The distribution of positive (and negative) charges on the various surfaces of the protein will also serve a critical role in determining the structure and activity of the protein. In particular, the distribution of positively-charged residues in an α-helical region of a protein can result in positive charges lying on one face of the helix or may result in the charged residues being concentrated in some particular portion of the molecule. An alternative distribution of positively charged residues is for them to project into the solvent in a radial orientation to the core of the protein. This orientation is predicted for several of the MiAMP2 homologues (data not shown). The spacing which is required for positioning of the residues on one face of the helix or the spacing required to accomplish a radial orientation from the core can easily be determined by one skilled in the art using a helical wheel plot with the sequence of interest. A helical wheel plot uses the fact that, in α-helices, each turn of the helix is composed of 3.6 residues on average. This number translates to 100° of rotational translation per residue making it possible to construct a plot showing the distribution of side chains in a helical region. Figure 8 shows how the spacing of charged residues can lead to most of the

positively charged side chains being localised on one face of the helix. It will be appreciated by one of skill in the art that positive charges are conferred by arginine and lysine residues.

In order for the protein to develop into a helix-turn-helix structure, it is also necessary to have particular residues that favor α-helix formation and that also favor a turn structure in the middle portion of the amino acid sequence (and disfavor a helical structure in the turn region). This can be accomplished by a proline residue or residues in the middle of the turn segment as seen with many of the MiAMP2 homologues. When proline is not present, glycine can also contribute to breaking a continuous helix structure, and inducing the formation of a turn at this position. In one case (i.e., TcAMP1), it appears that serine may be taking on this role. It will be appreciated that the residues in this region of the protein will usually favor the fomation of a turn structure; residues which fulfill this requirement include proline, glycine, serine, and aspartic acid; but, other residues are also allowed.

The DNA sequences reported here are an extremely powerful tool which can be used to obtain homologous genes from other species. Using the DNA sequences, one skilled in the art can design and synthesise oligonucleotide probes which can be used to screen cDNA libraries from other species of plants for the presence of genes encoding antimicrobial proteins homologous to the ones described here. This would simply involve construction of a cDNA library and subsequent screening of the library using as the oligonucleotide probe one or part of one of the sequences reported here (such as sequence ID. No. 2 or the PCR fragment described in Example 9). Other oligonucleotide sequences coding for proteins homologous to MiAMP2 can also be used for this purpose (e.g., DNA sequences corresponding to cotton and cocoa vicilins). Making and screening of a cDNA library can be carried out by purchasing a kit for said purpose (e.g., from Stratagene) or by following well established protocols described in available DNA cloning manuals (see *Current Protocols in Molecular Biology, supra*). It is relatively straight forward to construct libraries of various species and to specifically isolate vicilin homologues which are similar to the Macadamia, cotton, or cocoa vicilins by using a simple DNA hybridization technique to screen such libraries. Once cloned, these vicilin-related sequences can then be examined for the presence of MiAMP2-like subunits. Such subunits can easily be expressed in *E. coli* using the system described in Examples 10 and 11. Subsequently, these proteins can also be expressed in transgenic.

Genes, or fragments thereof, under the control of a constitutive or inducible promoter, can then be cloned into a biological system which allows expression of the protein encoded thereby. Transformation methods allowing for the protein to be expressed in a variety of systems are known. The protein can thus be expressed in any suitable system for the purpose of producing the protein for further use. Suitable hosts for the expression of the protein include *E. coli*, fungal cells, insect cells,

mammalian cells, and plants. Standard methods for expressing proteins in such hosts are described in a variety of texts including section 16 (Protein Expression) of *Current Protocols in Molecular Biology* (*supra*).

Plant cells can be transformed with DNA constructs of the invention according to a variety of known methods (*Agrobacterium*, Ti plasmids, electroporation, micro-injections, micro-projectile gun, and the like). DNA sequences encoding the *Macadamia integrifolia* antimicrobial protein subunits (i.e. fragments a, b, c, or d from the MiAMP2 clones) as well as DNA coding for other homologues can be used in conjunction with a DNA sequence encoding a preprotein from which the mature protein is produced. This preprotein can contain a native or synthetic signal peptide sequence which will target the protein to a particular cell compartment (e.g., the apoplast or the vacuole). These coding sequences can be ligated to a plant promoter sequence that will ensure strong expression in plant cells. This promoter sequence might ensure strong constitutive expression of the protein in most or all plant cells, it may be a promoter which ensures expression in specific tissues or cells that are susceptible to microbial infection and it may also be a promoter which ensures strong induction of expression during the infection process. These types of gene cassettes will also include a transcription termination and polyadenylation sequence 3' of the antimicrobial protein coding region to ensure efficient production and stabilisation of the mRNA encoding the antimicrobial proteins. It is possible that efficient expression of the antimicrobial proteins disclosed herein might be facilitated by inclusion of their individual DNA sequences into a sequence encoding a much larger protein which is processed *in planta* to produce one or more active MiAMP2-like fragments.

Gene cassettes encoding the MiAMP2 series antimicrobial proteins (i.e., MiAMP2a, b, c, or d; or all of the subunits together; or the entire MiAMP2 clone) or homologues of the MiAMP2 proteins as described above can then be expressed in plant cells using two common methods. Firstly, the gene cassettes can be ligated into binary vectors carrying: i) left and right border sequences that flank the T-DNA of the *Agrobacterium tumefaciens* Ti plasmid; ii) a suitable selectable marker gene for the selection of antibiotic resistant plant cells; iii) origins of replication that function in either *A. tumefaciens* or *Escherichia coli*; and iv) antibiotic resistance genes that allow selection of plasmid-carrying cells of *A. tumefaciens* and *E. coli*. This binary vector carrying the chimaeric MiAMP2 encoding gene can be introduced by either electroporation or triparental mating into *A. tumefaciens* strains carrying disarmed Ti plasmids such as strains LBA4404, GV3101, and AGL1 or into *A. rhizogenes* strains such as A4 or NCCP1885. These *Agrobacterium* strains can then be co-cultivated with suitable plant explants or intact plant tissue and the transformed plant cells and/or regenerants selected using antibiotic resistance.

A second method of gene transfer to plants can be achieved by direct insertion of the gene in target plant cells. For example, an MiAMP2-encoding gene cassette can be co-precipitated onto gold or tungsten particles along with a plasmid encoding a chimaeric gene for antibiotic resistance in plants. The tungsten particles can be accelerated using a fast flow of helium gas and the particles allowed to bombard a suitable plant tissue. This can be an embryogenic cell culture, a plant explant, a callus tissue or cell suspension or an intact meristem. Plants can be recovered using the antibiotic resistance gene for selection and antibodies used to detect plant cells expressing the MiAMP2 proteins or related fragments.

The expression of MiAMP2 proteins in the transgenic plants can be detected using either antibodies raised to the protein(s) or using antimicrobial bioassays. These and other related methods for the expression of MiAMP2 proteins or fragments thereof in plants are described in *Plant Molecular Biology* (2nd ed., edited by Gelvin, S.B. and Schilperoort, R.A., © 1994, published by Kluwer Academic Publishers, Dordrecht, The Netherlands)

Both monocotyledonous and dicotyledonous plants can be transformed and regenerated. Examples of genetically modified plants include maize, banana, peanut, field peas, sunflower, tomato, canola, tobacco, wheat, barley, oats, potato, soybeans, cotton, carnations, roses, sorghum. These, as well as other agricultural plants can be transformed with the antimicrobial genes such that they would exhibit a greater degree of resistance to pathogen attack. Alternatively, the proteins can be used for the control of diseases by topological application.

The invention also relates to application of antimicrobial protein in the control of pathogens of mammals, including humans. The protein can be used either in topological or intravenous applications for the control of microbial infections.

As indicated above in the description of the tenth embodiment, the invention includes within its scope the preparation of antimicrobial proteins based on the prototype MiAMP2 series of proteins. New sequences can be designed from the MiAMP2 amino acid sequences which substantially retain the distribution of positively charged residues relative to cysteine residues as found in the MiAMP2 proteins. The new sequence can be synthesised or expressed from a gene encoding the sequence in an appropriate host cell. Suitable methods for such procedures have been described above. Expression of the new protein in a genetically engineered cell will typically result in a product having a correct three-dimensional structure, including correctly formed disulphide linkages between cysteine residues. However, even if the protein is chemically synthesised, methods are known in the art for further processing of the protein to break undesireable disulfide bridges and form the bridges between the desired cysteine residues to give the desired three-dimensional structure should this be necessary.

Macadamia integrifolia antimicrobial proteins series number 2

As indicated above, a new series of potent antimicrobial proteins has been identified in the seeds of *Macadamia integrifolia*. The proteins collectivelly are called the MiAMP2 series of antimicrobial proteins (or MiAMP2 proteins) because they are all found on one large preproprotein which is processed into smaller subunits, each exhibiting antimicrobial activity; they represent the second class of antimicrobial proteins isolated from *Macadamia integrifolia*. Each protein fragment of the series has a characteristic pI value. MiAMP2a, b, c, and d subunits as shown in Figure 4 have predicted pI values of 4.4, 4.6, 11.5, and 11.6 respectively (predicted using raw sequence data without the His tag or cleavage sequences associated with expression of fragments in the vector pET16b), and contain two sets of CXXXC motifs which are important in stabilising the three-dimensional structure of the protein through the formation of disulfide bonds. Additionally, the proteins contain either an added set of aromatic (tyrosine/phenylalanine) residues or an added set of cysteine residues located at positions which would give more stability to the helix-turn-helix structure as described above and in Example 8.

The amino acid sequences of the MiAMP2 series of proteins share significant homology with fragments of previously described proteins in sequence databases (Swiss Prot and Non-redundant databases) searched using the BLASTP algorithm (*Altschul, S.F. et al.* [1990] *J. Mol. Biol.* 215:403). In particular, MiAMP2a, b, c and d sequences exhibit significant similarity with regions of cocoa vicilin and cotton vicilin (as seen in Figure 6). Some similarity is also seen with fragments from other seed storage proteins of peanut (Burks, A. W. *et al.* [1995] *J. Clin. Invest.* 96 (4), 1715-1721), maize (Belanger, F. C. and Kriz, A. L.[1991] *Genetics* 129 (3), 863-872), barley (Heck, G. R. *et al.* [1993] *Mol. Gen. Genet.* 239 (1-2), 209-218), and soybean (Sebastiani, F. L. *et al.* [1990] *Plant Mol. Biol.* 15 (1), 197-201). Although, in some cases the homology is not extremely high (for example, 18% identity between MiAMP2a and cotton subunit 1; see Figure 4), the spacing of the main four cysteine residues is conserved in all subunits and homologues. In addition, both cotton and cocoa vicilin-derived subunits retain the conserved tyrosine or phenylalanine residues as additional stabilisers of the tertiary structure. The cotton and cocoa vicilins with 525 and 590 amino acids, respectively, are much larger proteins than MiAMP2c (47 amino acids) (see Figures 4 and 6). Although MiAMP2 subunits also share some homology with MBP-1 antimicrobial protein from maize (Duvick, J.P. *et al.* (1992) *J Biol Chem* 267:18814-20) the number of residues between the CXXXC motifs is 13 which puts MBP-1 outside the specifications for the spacing given here in this application. MBP-1 is also a smaller protein (33 amino acids), overall, than the sequences claimed here and there is no evidence available the MBP-1 is derived from a larger seed storage protein other than some similarity with a portion of miaze globulin protein. However, MBP-1 cannot be derived

from from the maize globulin since maize globulin contains 10 residues between the two CXXXC motifs while MBP-1 contains 13. The alignments in Figures 4 and 6 show the similarity in cysteine spacing between MiAMP2 subunits and the cocoa and cotton vicilin-derived molecules. The cysteine and the aromatic tyrosine/phenylalanine residues in Figures 4 and 6 are highlighted with bold underlined text. Figure 4 also shows the alignment of additional proteins which can be expressed in liquid culture and shown to exhibit antimicrobial activity.

All of the MiAMP2 homologues that have been tested exhibit antifungal activity. MiAMP2 homologues show very significant inhibition of fungal growth at concentrations as low as 2 μg/ml for some of the pathogens/microbes against which the proteins were tested. Thus they can be used to provide protection against several plant diseases. MiAMP2 homologues can be used as fungicides or antibiotics by application to plant parts. The proteins can also be used to inhibit growth of pathogens by expressing them in transgenic plants. The proteins can also be used for the control of human pathogens by topological application or intravenous injection. One characteristic of the proteins is that inhibition of some microbes is suppressed by the presence of Ca^{2+} (1 mM). An example of this effect is provided for MiAMP2c subunit in Table 1.

Some of the MiAMP2 proteins and homologues could also function as insect control agents. Since some of the proteins are extremely basic (e.g., pI > 11.5 for MiAMP2c and d subunits), they would maintain a strong net-positive charge even in the highly alkaline environment of an insect gut. This strong net-positive charge would enable it to interact with negatively charged structures within the gut. This interaction may lead to inefficient feeding, slowing of growth, and possibly death of the insect pest.

Non-limiting examples of the invention follow.

Example 1

Extraction of Basic Protein from *Macadamia integrifolia* Seeds

Twenty five kilograms of Mi nuts (purchased from the Macadamia Nut Factory, Queensland, Australia) were ground in a food processor (The Big Oscar, Sunbeam) and the resulting meal was extracted for 2-4 hours at 4°C with 50 L of an ice-cold extraction buffer containing 10 mM NaH_2PO_4, 15 mM Na_2HPO_4, 100 mM KCl, 2 mM EDTA, 0.75% polyvinylpolypyrolidone, and 0.5 mM phenylmethylsulfonyl fluoride (PMSF). The resulting homogenate was run through a kitchen strainer to remove larger particulate material and then further clarified by centrifugation (4000 rpm for 15 min) in a large capacity centrifuge. Solid ammonium sulphate was added to the supernatant to obtain 30% relative saturation and the precipitate allowed to form overnight with stirring at 4°C. Following centrifugation at 4000 rpm for 30 min, the supernatant was taken and ammonium sulphate added to achieve 70% relative saturation. The solution was allowed to precipitate overnight and then

centrifuged at 4000 rpm for 30 min in order to collect the precipitated protein fraction. The precipitated protein was resuspended in a minimal volume of extraction buffer and centrifuged once again (13,000 rpm x 30 min) to remove the any insoluble material yet remaining. After dialysis (10 mM ethanolamine pH 9.0, 2 mM EDTA and 1 mM PMSF) to remove residual ammonium sulphate, the protein solution was passed through a Q-Sepharose Fast Flow column (5 x 12 cm) previously equilibrated with 10 mM ethanolamine (pH 9), 2 mM in EDTA). The collected flowthrough from this column represents the basic (pI >9) protein fraction of the seeds. This fraction was further purified as described in Example 3.

Example 2

Antifungal and Antibacterial Activity Assays

In general, bioassays to assess antifungal and antibacterial activity were carried out in 96-well microtitre plates. Typically, the test organism was suspended in a synthetic growth medium consisting of K_2HPO_4 (2.5 mM), $MgSO_4$ (50 μM), $CaCl_2$ (50 μM), $FeSO_4$ (5 μM), $CoCl_2$ (0.1 μM), $CuSO_4$ (0.1 μM), Na_2MoO_4 (2 μM), H_3BO_3 (0.5 μM), KI (0.1 μM), $ZnSO_4$ (0.5 μM), $MnSO_4$ (0.1 μM), glucose (10 g/L), asparagine (1 g/L), methionine (20 mg/L), myo-inositol (2 mg/L), biotin (0.2 mg/L), thiamine-HCl (1 mg/L) and pyridoxine-HCL (0.2 mg/L). The test organism consisted of bacterial cells, fungal spores (50,000 spores/ml) or fungal mycelial fragments (produced by blending a hyphal mass from a culture of the fungus to be tested and then filtering through a fine mesh to remove larger hyphal masses). Fifty microlitres of the test organism suspended in medium was placed into each well of the microtitre plate. A further 50 μl of the test antimicrobial solution was added to appropriate wells. To deal with well-to-well variability in the bioassay, 4 replicates of each test solution were done. Sixteen wells from each 96-well plate were used as controls for comparison with the test solutions.

Unless otherwise stated, incubation was at 25°C for 48 hours. All fungi including yeast were grown at 25°C. *E. coli* were grown at 37°C and other bacteria were bioassayed at 28°C. Percent growth inhibition was measured by following the absorbance at 600 nm of growing cultures over various time intervals and is defined as 100 times the ratio of the average change in absorbance in the control wells minus the change in absorbance in the test well divided by the average change in absorbance at 600 nm for the control wells (i.e., [(avg change in control wells - change in test well) / (avg change in control wells)] x 100). Typically, measurements were taken at 24 hour intervals and the period from 24-48 hours was used for %Inhibition measurements.

Example 3

Purification of antimicrobial protein from *Macadamia integrifolia* basic protein fraction

The starting material for the isolation of the Mi antimicrobial protein was the basic fraction extracted from the mature seeds as described above in Example 1. This protein was further purified by cation exchange chromatography as shown in Figure 1.

About 4 g of the basic protein fraction dissolved in 20 mM sodium succinate (pH 4) was applied to an S-Sepharose High Performance column (5 X 60 cm) (Pharmacia) previously equilibrated with the succinate buffer. The column was eluted at 17 ml/min with a linear gradient of 20 L from 0 to 2 M NaCl in 20 mM sodium succinate (pH 4). The eluate was monitored for protein by on-line measurement of the absorbance at 280 nm and collected in 200 ml fractions. Portions of each fraction were subsequently tested in the antifungal activity assay against *Phytopthora cryptogea* at a concentration of 100 μg/ml in the presence and absence of 1 mM Ca^{2+}. Results of bioassays are included in Figures 1a and 1b where the elution gradient is shown as a solid line and the shaded bars represent %Inhibition. The Figure 1a assays were conducted without added Ca^{2+} while 1 mM Ca^{2+} was included in the Figure 1b assays. Fractionation yielded a number of unresolved peaks eluting between 0.05 and 2 M NaCl. A peak eluting at about 16 hours into the separation (fractions 90-92) showed significant antimicrobial activity.

Fractions showing significant antimicrobial activity were further purified by reversed-phase chromatography. Aliquots of fractions 90-92 were loaded onto a Pep-S (C_2/C_{18}), column (25 x 0.93 cm) (Pharmacia) equilibrated with 95% H_2O/5% MeCN/0.1% TFA (=100%A). The column was eluted at 3 ml/min with a 240 ml linear gradient (80 min) from 100%A to 100%B (=5% H_2O/95% MeCN/0.1% TFA). Individual peaks were collected, vacuum dried three times in order to remove traces of TFA, and subsequently resuspended in 500 microlitres of milli-Q water (Millipore Corporation water purification system) for use in bioassays as described in Example 2. Figure 2 shows the HPLC profile of purified fraction 92 from the cation-exchange separation shown in Figures 1 and 2. Protein elution was monitored at 214 nm. The acetonitrile gradient is shown by the straight line. Individual peaks were bioassayed for antimicrobial activity: the bars in Figure 3 show the inhibition corresponding to 15 μg/ml of material from each of the fractions. The active protein elutes at approximately 27 min (~30% MeCN/0.1%TFA) and is called MiAMP2c.

Example 4

Purity of Isolated MiAMP2c

The purity of the isolated antimicrobial protein was verified by native SDS-PAGE followed by staining with coomassic blue protein staining solution. Electrophoresis was performed on a 10-20% tricine gradient gel (Novex) as per the manufacturers recommendations (100 V, 1-2 hour separation

time). Under these conditions the purified MiAMP2c migrates as a single discrete band (<10 kDa in size). The detection of a single major band in the SDS-PAGE analysis together with single peaks eluting in the cation-exchange and reversed-phase separations (not shown), gives strong indication that the MiAMP2c preparation is greater than 95% pure and therefore the activity of the preparation was almost certainly due to the MiAMP2c alone and not to a minor contaminating component. A clean signal in mass spectrometric analysis (Example 5 below) also supports this conclusion.

Example 5

Mass Spectroscopic Analysis of MiAMP2c

Purified MiAMP2c was submitted for mass spectroscopic analysis. Approximately 1 μg of protein in solution was used for testing. Analysis showed the protein to have a molecular weight of 6216.8 Da ± 2 Da. Additionally, the protein was subjected to reduction of disulfide bonds with dithiothreitol and alkylation with 4-vinylpyridine. The product of this reduction/alkylation was then submitted for mass spectroscopic analysis and was shown to have gained 427 mass units (i.e. molecular weight was increased by approximately 4 X 106 Da). The gain in mass indicated that four 4-vinylpyridine groups had reacted with the reduced protein, demonstrating that the protein contains a total of 4 cysteine residues. The cysteine content has also been subsequently confirmed through amino acid sequencing.

Example 6

Amino Acid Sequence of MiAMP2c Protein

Approximately 1 μg of the pure protein which had been reduced and alkylated was subjected to Automated Edman degradation N-terminal sequencing. In the first sequencing run, the sequence of the first 39 residues was determined. Subsequently, approximately 1 mg of MiAMP2c was reacted with Cyanogen Bromide which cleaved the protein on the C-terminal side of Methionine-26. The C-terminal fragment generated by the cleavage reaction was purified by reversed-phase HPLC and sequenced, yielding the remaining sequence of MiAMP2c (i.e. residues 27-47). The full amino acid sequence is RQRDP QQQYE QCQER CQRHE TEPRH MQTCQ QRCER RYEKE KRKQQ KR and represents amino acids 118 to 164 of clone 3 from Example 9 (see Figure 6 and SEQUENCE ID NO: 5). In the figure, cysteine residues are in bold type and underlined to facilitate recognition of the spacing patterns. Depending on the number of disulfide bonds that are formed, the protein mass will range from 6215.6 to 6219.6 Da. This is in close agreement with the mass of 6216.8 ± 2 Da obtained by mass spectrometric analysis (Example 5). The measured mass closely approximates the predicted mass of MiAMP2c in a two-disulfide form as is expected to be the case.

Example 7

Synthetic DNA Sequence Coding for MiAMP2c with a leader peptide

Using standard codon tables it is possible to reverse-translate the protein sequences to obtain DNA sequences that will code for the antimicrobial proteins. The software program MacVector 4.5.3 was used to enter the protein sequence and obtain a degenerate nucleotide sequence. A codon usage table for tobacco was referenced in order to pick codons that would be adequately represented in tobacco for purposes of obtaining high expression in this test plant. A 30 amino-acid leader peptide was also designed to ensure efficient processing of the signal peptide and secretion of the peptide extracellularly. For this purpose, the method of Von Hiejne was used to evaluate a series of possible leader sequences for probability of cleavage at the correct position [Von Hiejne, G.(1986) *Nucleic Acids Research* 14(11): 4683-4690]. In particular, the amino acid sequence MAWFH VSVCN AVFVV IIIIM LLMFV PVVRG (Sequence ID. No. 11) was found to give an optimal probability of correct processing of the signal peptide immediately following the G (Gly) of this leader sequence. A 5' untranslated region from tobacco mosaic virus was also added to this synthetic gene to promote higher translational efficiency [Dowson, M.J., *et al.* (1994) *Plant Mol. Biol. Rep.* 12(4):347-357]. The synthetic gene also contains restriction sites at the 5' and 3' ends and immediately 5' of the start ATG for efficient cloning and subcloning procedures. Figure 5 shows a synthetic DNA sequence suitable for use in plant expression experiments. In this Figure, the arrow shows where translation is initiated and the triangular symbol indicates the point of cleavage of the signal peptide.

Example 8

Structure prediction of MiAMP2c Protein

Using sequence analysis algorithms, putative secondary structure motifs can be assigned to the protein. Five different algorithms were used to predict whether α-helices, β-sheets, or turns can occur in the MiAMP2c protein (Figure 4). Methods were obtained from the following sources: DPM method, Deleage, G., and Roux, B. (1987) *Prot. Eng.* 1:289-294; SOPMA method, Geourjon, C., and Deleage, G. (1994) *Prot. Eng.* 7:157-164; Gibrat method, Gibrat, J.F., Garnier, J., and Robson, B.(1987) *J.Mol.Biol.* 198:425-443; Levin method, Levin, J.M., Robson, B., and Garnier, J. (1986) *FEBS Lett.* 205:303-308; and PhD method, Rost, B., And Sander, C. (1994) *Proteins* 19:55-72. Figure 7 shows the predicted locations of α-helices, β-sheets and turns. The following symbols have been used in Figure 7: C, coil (unstructured); H, alpha helix; E, β- sheet; and S, turn. Underlined residues are those which were predicted to exhibit an α-helical structure by at least 2 separate structure prediction methods; these are represented as helices in Figure 8.

It is clear from the secondary structure predictions that the protein is highly α-helical. While secondary structure prediction is often difficult and inaccurate, this particular prediction gives a clear indication of the structure of the protein. Examination of the secondary-structure predictions show a clear preponderance of two α-helical regions broken by a stretch of about 5-8 residues. This is highly suggestive of a helix-turn-helix motif.

Helical wheel analysis of the MiAMP2c amino acid sequence shows that cysteine residues with a CXXXC spacing will be aligned on one face of the helix in which they are located. Since the cysteines are involved in disulfide bond formation, the cysteine side chains in one helix must form covalent bonds with the cysteine side chains located on the other helical segment. When the helical segments are arranged in such a way as to bring the cysteine side chains from each respective helix into proximity with the other cysteine side chains, the resulting three-dimensional structure is shown in Figure 8. This structure exhibits a remarkable distribution of positively charge residues on one face of the protein comprised of two helices held together by two disulfide bonds. Figure 8 shows how the spacing of positively charged residues in helical regions of this molecule will cause these side chains to lie on one face of the helix. The positively charged residues are the dark side chains outlined in black. Other dark side chains represent acidic residues. A proline residue (grey colour marked with a 'P') is located at the extreme left end of the molecule in the turn region. Solid black lines show where disulfide bonds connect the two helices. The dotted line shows where the two aromatic hydrophobic residues interact to add stability to the helix-turn-helix structure.

This helix-turn-helix structure will be adopted by all MiAMP2 homologues containing the same cysteine spacing and residues with helix and turn-forming propensities. Other MiAMP2 fragment sequences can be superimposed onto the global structure shown in figure 8. The overall structure will remain essentially the same but the charge distribution will vary according to the sequences involved. In the case of MiAMP2b, the dotted line would represent an added disulfide bridge instead of a hydrophobic interaction.

Example 9

cDNA cloning of genes corresponding to MiAMP2c

PCR Amplification of a genomic fragment of the MiAMP2c gene

Using the reverse-translated nucleotide sequences, degenerate primers were made for use in PCR reactions with genomic DNA from Macadamia. Primer JPM17 sequence was 5' CAG CAG CAG TAT GAG CAG TG 3' and primer JPM20 degenerate sequence was 5' TTT TTC GTA (T/T)C(T/G) (G/T)C(T/G) TTC GCA 3' (SEQ ID NOS: 12 and 13). Primers JPM17 and JPM20 were used in PCR amplifications carried out for 30 cycles with 30 sec at 95°C, 1 min at 50°C, and 1 min at 72°C. PCR products with sizes close to those which were expected were directly sequenced

(ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit from Perkin Elmer Corporation) after excising DNA bands from agarose gels and purifying them using a Qiagen DNA clean-up kit. Using this approach, it was possible to amplify a fragment of DNA of approximately 100 bp. Direct sequencing of this nucleotide fragment yielded the nucleotide sequence corresponding to a portion of the amino acid sequence of the antimicrobial protein MiAMP2c (amino acids 7-39 of Figure 4). The partial nucleotide sequence obtained from the above-mentioned fragment excluding the primer sequences was 5' TCA GAA GCG CTG CCA ACG GCG CGA GAC AGA GCC ACG ACA CAT GCA AAT TTG TCA ACA ACG C 3' (corresponding to base pairs 264 to 324 in SEQ ID NO: 6). This sequence can be used for a variety of purposes including screening of cDNA and genomic libraries for clones of MiAMP2 homologues or design of specific primers for PCR amplification reactions.

Messenger RNA isolation from Macadamia nut kernels

Fifty-eight grams of Macadamia nut kernels were ground to powder under liquid nitrogen using a mortar and pestel. RNA from ground material was then purified using a Guanidine thiocyanate/Cesium chloride technique (*Current Protocols in Molecular Biology, supra*). Using this method approximately 5 mg of total RNA was isolated. Messenger RNA was then purified from total RNA using a spun column mRNA purification kit (Pharmacia).

cDNA library construction

A cDNA library was constructed in a lambda ZAP vector using a library kit from Stratagene. A total of 6 reactions were performed using 25 micrograms of messenger RNA. First and second strand cDNA synthesis was performed using MMLV Reverse transcriptase and DNA Polymerase I, respectively. After blunting the cDNA with *Pfu* DNA Polymerase, *Eco* RI linker adapters were ligated to the DNA. DNA was then kinased using T4 polynucleotide kinase and the DNA subsequently digested with *Xho* I restriction endonuclease. At this point cDNA material was fractionated according to size using a sephacryl-S500 column supplied with the kit. DNA was then ligated into the lambda ZAP vector. The vector containing ligated insert was then packaged into lambda phage (Gigapack III packaging extract from Stratagene).

Screening of library

The library constructed above was then plated and screened in XL1-blue *E.coli* bacterial lawns growing in top agarose. Plaques containing individual clones were isolated by lifting onto Hybond N+ membranes (Amersham LIFE SCIENCE), hybridizing to a radiolabeled version of the genomic DNA fragment amplified above, imaging of the blot, and picking of possitive clones for the next round of screeing. After secondary and tertiary screening, plaques were sufficiently isolated to allow

picking of single clones. Several clones were obtained, and subsequently the pBK-CMV vector portion from the larger lambda vector was excised.

Sequence of MiAMP2c cDNA clones

Vectors (pBK-CMV) containing putative MiAMP2c clones were sequenced to obtain the DNA sequence of the cloned inserts. Seven clones were partially sequenced and an additional three clones were fully sequenced (see SEQ ID NOS: 2, 4 and 6 for DNA sequences of the macadamia clones). Translation of the DNA sequences showed that the full length clones encoded highly similar proteins of 666 amino acids. Figure 6 shows that these proteins have substantial similarity to vicilin seed-strorage proteins from cocoa and cotton. Stars show positions of conserved identities and dots show positions of conserved similarities. Examination of the protein sequences revealed that the exact MiAMP2c sequence is found within the translated protein sequence of clone 3 at amino acid positions 118 to 164 (see Figure 6); clones 1 and 2 contained sequences differing from MiAMP2c by 2 residues and 3 residues, respectively, out of 47 amino acids total in the MiAMP2c sequence.

The translation products of the full-length clones (i.e., clones 1 and 2) consist of a short signal peptide from residues 1 to 28, a hydrophilic region from residues 29 to ~246, and then two segments stretching from residues ~246 to 666 with a stretch of acidic residues separating them at positions 542-546.

Significantly, the hydrophilic region containing the sequence for MiAMP2c, also contains 3 additional segments which are very similar to MiAMP2 (termed MiAMP2a, b and d). These 4 segments (found between residues 28 and ~246) are separated by stretches in which approximately four out of five residues are acidic (usually glutamic acid). These acidic stretches occur at positions 64-68, 111-115, 171-174, and 241-246 and appear to delineate processing sites for cleavage of the 666-residue preproprotein into smaller functional fragments (acidic stretches delineating cleavage sites are shown as bold characters in Figure 6). All four MiAMP2-like segments of the protein contain 2 doublets of cysteine residues separated by 10-12 residues to give the following pattern C-X-X-X-C-(10-12X)-C-X-X-X-C where X is any amino acid, and C is cysteine. All four segments are expected to form helix-turn-helix motifs as decribed in Example 8 above. It is clear that the cysteines in these locations will form disulfide bridges that stabilize the structure of the proteins by holding the two helical portions together.

The predicted helix-turn-helix motifs can be further stabilized in several ways. The first method of stabilization is exemplified in segments 1 and 3 (i.e., residues 29-63 and 118-170, respectively, of the 666-residue Macadamia vicilin-like protein). These segments is the are stabilized by a hydrophobic ring-stacking interaction between two aromatic residues (one on each α-helical segment); this is normally accomplished with tyrosine residues but phenylalanine is also

used. As with the cysteine residues, the location of these aromatic residues in the predicted α-helical segments is critical if they are to offer stabilization to the helix-turn-helix structure. In segments 1 and 3, the aromatic residues are 2 and 3 residues removed from the cysteine doublets as shown here: Z-X-X-C-X-X-X-C-(10-12X)-C-X-X-X-C-X-X-X-Z where C is cysteine and Z is usually tyrosine but can be substituted with phenylalanine as is done in segment 1.

The second way to stabilize the helix-turn-helix fragment is by using an added disulfide bridge as seen in fragment 2 (residues 71-110). This is accomplished by placing additional cysteine residues 2 and 3 residues removed from the cysteine doublets as shown here: nX-C-X-X-C-X-X-X-C-(10-12X)-C-X-X-X-C-X-X-X-C-nX. This is the only report that the inventors know of where a helix-turn-helix domain in an antimicrobial protein is stabilized by three disulfide bridges. While segment 4 (residues 175-241) does not contain the extra disulfide bridge or the hydrophobic ring-stacking stabilization, it is probably stabilized by means of weaker ionic and or hydrogen bonding interactions.

Example 10

Vectors for liquid culture expression of MiAMP2 and homologues

PCR primers flanking the nucleotide region coding for MiAMP2c were engineered to contain restriction sites for *Nde* I and *Bam* HI (corresponding to the 5' and 3' ends of the coding region, respectively; Primer JPM31 sequence: 5' A CAC CAT ATG CGA CAA CGT GAT CC 3'; Primer JPM32 sequence: 3' C GTT GTT TTC TCT ATT CCT AGG GTT G 5', SEQ ID NOS: 14 and 15). These primers were then used to amplify the coding region of MiAMP2c DNA. The PCR product from this amplification was then digested with *Nde* I and *Bam* HI and ligated into a pET17b vector (Novagen / Studier, F. W. *et al.* [1986] *J. Mol. Biol.* 189:113) with the coding region in-frame to produce the vector pET17-MiAMP2c.

A similar approach to the one above was used to construct vectors carrying the coding sequences of MiAMP2c homologues (i.e. MiAMP2a, b, and d as well as Tc AMP1, and TcAMP2). To construct the expression vectors for fragments a, b and d in MiAMP2 clone 1, specific PCR primers incorporating the *Nde* I and *Bam* HI sites were designed to amplify the fragments of interest. The products were then digested with the appropriate restriction enzymes and ligated into the *Nde* I/*Bam* HI sites of a pET16b vector [Novagen] containing a His tag and a Factor Xa cleavage site (amino acid sequence MGHHH HHHHH HHSSG HIEGR HM, SEQ ID NO: 16). The protein products expressed from the pET16b vector is a fusion to the antimicrobial protein. The coding sequences for MiAMP2-like subunits from cocoa (Figure 4, TcAMP1 and TcAMP2) were obtained from the published DNA sequence of the cocoa vicilin gene (Spencer, M. E. and Hodge R. [1992] *Planta* 186:567-576). Two MiAMP2-like fragments within the cocoa vicilin gene were located at

the 5' end (corresponding to the residues shown in Figure 4), and two sets of complimentary oligonucleotides corresponding to the desired coding sequences were designed. The complimentary oligonucleotides (90 to ~100 bases) corresponding to each cocoa subunit contained a 20bp overlap and also contained the *Nde* I and *Bam* HI restriction endonuclease cut sites.

For TcAMP, the following nucleotides were synthesised:

TcAMP1 forward oligo 5' GGGAATTCCA TATGTATGAG CGTGATCCTC GACAGCAATA CGAGCAATGC CAGAGGCGAT GCGAGTCGGA AGCGACTGAA GAAAGGGAGC 3';

TcAMP1 reverse oligo 5' GAAGCGACTG AAGAAAGGGA GCAAGAGCAG TGTGAACAAC GCTGTGAAAG GGAGTACAAG GAGCAGCAGA GACAGCAATA GGGATCCACA C 3'.

For TcAMP2, the following oligonucleotides were used:

TcAMP2 forward oligo 5' GGGAATTCCA TATGCTTCAA AGGCAATACC AGCAATGTCA AGGGCGTTGT CAAGAGCAAC AACAGGGGCA GAGAGAGCAG CAGCAGTGCC AGAGAAAATG C 3';

TcAMP2 reverse oligo 5' GTGTGGATCC CTAGCTCCTA TTTTTTTTGT GATTATGGTA ATTCTCGTGC TCGCCTCTCT CTTGTTCCTT ATATTGCTCC CAGCATTTTC TCTGGCACTG CT 3'.

The oligonucleotide sets were added to individual PCR amplification reactions in order make individual PCR fragments containing the desired coding region. Since initial PCR amplifications gave fuzzy bands, reamplification of the original products was carried out using new 20mer primers (complimentary to the 5'ends of the forward and reverse oligonucleotides shown above) designed to amplify the entire coding region of the cocoa subunits. Once amplified, the PCR products were restriction digested with the appropriate enzymes and ligated into the vector pET16b as above. This procedure was carried out for both cocoa fragments with similarities to MiAMP2c (shown in Figure 4).

Example 11

Expression in *E.coli* and purification of MiAMP2c and homologues

Starter cultures (50 ml) of *E.coli* strain BL21 (Grodberg, J. [1988] *J. Bacteriol.* 170:1245) transformed with the appropriate pET construct (Example 10) were added to 500ml of NZCYM media (*Current Protocols in Molecular Biology, supra*) and cultured to an optical density of 0.6 (600 nm) and induced with the addition of 0.4 or 1.0 mM IPTG depending on whether pET17b

(containing a T7 promoter) or pET16b (containing a His tag fusion and a T7 promoter/lac operator) vector was being used. After cells were induced, cultures were allowed to grow for 4 hours before harvesting. Aliquots of the growing cultures were removed at timed intervals and protein extracts run on an SDS-PAGE gel to follow the expression levels of MiAMP2 and homologues in the cultures. Fragments being expressed with a Histidine tag (i.e., in the pET16b vector), were harvested by centrifuging induced cell cultures at 5000g for 10 minutes. Cell pellets were resuspended and broken by stirring for one hour in 6 M Guanidine-HCl, buffered with 100 mM sodium phosphate and 10 mM Tris at pH 8.0. Broken cell suspensions were centrifuged at 10,000g for 20-30 minutes to settle the cellular debris. Supernatants were removed to fresh tubes and 500 mg of Ni-NTA fast flow resin (Qiagen) was added to each supternatant. After gentle mixing at 4°C for 30-60 minutes, the suspension was loaded into a small column, rinsed two times with 8 M Urea (pH 8.0 and then pH 6.3) and subsequently, the protein was eluted using 8 M Urea pH 4.5. Protein fractions thus obtained were substantially pure but were further purified using an 9.3 x 250 mm C2/C18 reverse phase column (Pharmacia) and 75 minute gradient from 5% to 50% acetonitrile (0.1% TFA) flowing at 3 ml/min (data not shown).

All of the MiAMP2c homologues (except MiAMP2c which was expressed in pET17b) were expressed in the pET16b vector containing the Histidine tag. While induction of the MiAMP2c culture proceded as above, the rest of the purification was somewhat different. In this case, MiAMP2c-expressing cells were harvested by centrifugation but were then resuspended in phosphate buffer (100 mM, pH 7.0 containing 10 mM EDTA and 1 mM PMSF) and broken open using a French press instrument. Cellular debris containing MiAMP2c inclusion bodies was solubilized using a 6 M Guanidine-HCl, 10 mM MES pH 6.0 buffer. Soluble material was then recovered after centrifugation to remove insoluble debris remaining from the solubilization step. Guanidine-HCl soluble material was then dialyzed against 10 mM MES pH 6.0 containing PMSF (1 mM) and EDTA (10 mM). Cation-exchange fractionation was carried out as described in Example 3 except on a smaller scale after the dialysis step. Subsequently, the major eluting protein from the cation-exchange column, which was MiAMP2c, was then further purified using reverse phase HPLC as described in Example 3.

Figure 9 shows the SDS-PAGE gel analysis of the various purification stages obtained following induction with IPTG and subsequent purification of expressed proteins. Samples analysed during the TcAMP1 purification were are as follows: lane 1, molecular weight markers; lane 2, Ni-NTA non-binding fraction; lane 3, rinse of Ni-NTA resin with pH 8 urea; lane 4, rinse of Ni-NTA resin with pH 6.3 urea; lane 5, elution of TcAMP1 with pH 4.5 urea; and lane 6, second elution of TcAMP1 with pH 4.5 urea. TcAMP2 was purified in a similar manner and was also subjected to

reverse-phase HPLC to further purify the fraction eluting from the Ni-NTA resin. Figure 10 shows the reverse phase purification of cocoa subunit number 2 (TcAMP2).

SDS-PAGE gel analysis of the MiAMP2a, b, and d fragment purification is shown in the second panel of Figure 9. Lane contents are as follows: lane 1, molecular weight markers; lane 2, MiAMP2a pre-induced cellular extractp; lane 3, MiAMP2a IPTG induced cellular extract; lane 4, MiAMP2a Ni-NTA non-binding fraction; lane 5, MiAMP2a elution from Ni-NTA; lane 6, MiAMP2b pre-induced cellular extract; lane 7, MiAMP2b IPTG induced cellular extract; lane 8, MiAMP2b Ni-NTA non-binding fraction; lane 9, MiAMP2b elution from Ni-NTA; lane 10, MiAMP2d pre-induced cellular extract; lane 11, MiAMP2d IPTG induced cellular extract; lane 12, MiAMP2d Ni-NTA non-binding fraction; and lane 13, MiAMP2d elution from Ni-NTA.

Using the vectors described in Example 10, MiAMP2c, and 5 homologues (i.e., MiAMP2a, MiAMP2b, MiAMP2d, TcAMP1 and TcAMP2) were all expressed, purified and tested for antimicrobial activity. The approach taken above can be applied to all of the antimicrobial fragments described in Figure 4. Purified fragments can then be tested for specific inhibition agains microbial pathogens of interest.

Example 12

Detection of MiAMP2 homologues in other species using antibodies raised to MiAMP2c

Rabbits were immunised intramuscularly according to standard protocols with MiAMP2 conjugated to diphtheria toxoid suspended in Fruends incomplete adjuvent. Serum was harvested from the animals at regular intervals after giving the animal added doses of MiAMP2 adjuvent to boost the immune response. Approximately 100 ml of serum were collected and used for screening of crude extracts obtained from several plant seeds. One hundred gram quantities of seeds were ground and extracted to obtain a crude extract as in Example 1. Aliquots of protein were separated on SDS-PAGE gels and the gels were then blotted onto nitrocellulose membrane for subsequent detection of antibody reacting proteins. The membranes were incubated with MiAMP2c rabbit primary antibodies, washed and then incubated with alkaline phosphatase-conjugated goat anti-rabbit IgG for colorimetric detection of antigenic bands using the chemical 5-bromo-4-chloro-3-indolyl phosphate / nitroblue tetrazolium substrate system (Schleicher and Schuell). Figure 11 shows that various other species contain immunologically-related proteins of similar size to MiAMP2c. Lanes 1-15 contain the extracts from the following species: 1) *Stenocarpus sinuatus*, 2) *Stenocarpus sinuatus*(1/10 loading) , 3) *Restio tremulus*, 4) *Mesomalaena tetragona*, 5) *Nitraria billardieri*, 6) *Petrophile canescens*, 7) *Synaphae acutiloba*, 8) *Dryandra formosa*, 9) *Lambertia inermis*, 10) *Stirlingia latifolia*, 11) *Xylomelum angustifolium*, 12) *Conospermum bracteosum*, 13) *Conospermum triplinernium*, 14) Molecular weight marker, 15) *Macacamia integrifolia* pure MiAMP2c. Lanes 1-

13 contain a variety of species, some of which show the presence of antigenically related proteins of a similar size to MiAMP2c. Other bands exhibiting higher molecular weights probably represent the larger precursor seed storage proteins from which the antimicrobial proteins are derived. Antigenically-related proteins can be seen in lanes 1, 2, 4, 6, 7, 8, 9, and 11-13.

Bioassays were also performed using crude extracts from various Proteaceae species. Specifically, extracts from *Banksia robur*, *Banksia canei*, *Hakea gibbosa*, *Stenocarpus sinuatus*, and *Stirlingia latifolia* have all been shown to exhibit antimicrobial activity. This activity may derive from MiAMP2 homologues since these species are related to Macadamia.

Example 13

Purification of MiAMP2c homologues in another species using antibodies raised to MiAMP2c

Based on the detection of immunologically related proteins in other species of the family Proteaceae and the presence of antimicrobial activity in crude extracts, *Stenocarpus sinuatis* was chosen for a large scale fractionation experiment in an attempt to isolate MiAMP2c homologues. Five kg of *S.sinuatus* seed was frozen in liquid nitrogen and ground in a food processor (Big Oscaar Sunbeam). The ground seed was immediately placed into 12 L of 50 mM H_2SO_4 extraction buffer and extracted at 4°C for 1 hour with stirring. The slurry was then centrifuged for 20 min at 10,000 g to remove particulate matter. The supernatant was then adjusted to pH 9 using a 50mM ammonia solution. PMSF and EDTA were added to final concentrations of 1 and 10 mM respectively.

The crude protein extract was applied to an anion exchange column (Amberlite IRA-938, Rohm and Haas) (3cmx90cm) equilibrated with 50 mM NH_4Ac pH 9.0 at a flow rate of 40 ml/min. The unbound protein comprising the basic protein fraction was collected and used in the subsequent purification steps.

The basic protein fraction was adjusted to pH 5.5 with acetic acid and then applied at 10 ml/minute over 12 h to a SP-Sepharose Fast Flow (Pharmacia) Column (5cm x 60cm) pre-equilibrated with 25mM ammonium acetate. The column was then washed for 3.5 h with 25 mM Acetate pH 5.5. Elution of bound proteins was achieved by applying a linear gradient of NH_4Ac from 25 mM to 2.0 M (pH 5.5) at 10 ml/min over 10 h. Absorbance of the eluate was observed at 280 nm and 100 ml fractions collected (see Figure 12).

Cation-exchange fractions that cross-reacted with the antiserum (fractions 14-28, Figure 12) were then further purified by reverse phase chromatography. Cross-reacting fractions were loaded onto a 7 µm C18 reverse phase column (Brownlee) equilibrated with 90% H_20, 10% acetonitrile and 0.1% Trifluoroacetic acid (TFA)(=100%A). Bound proteins were eluted with a linear gradient from 100%A to 100%B (5% H20, 95% acetonitrile, 0.08% TFA). The absorbance of the eluted proteins was monitored at 214nm and 280nm. The eluted proteins were dried under vacuum and resuspended

in water three times to remove traces of TFA from the samples. Reverse phase protein elution fractions 20 to 61 were analysed by pooling 2 adjacent fractions and performing a western blot analysis (see Figure 13). Fractions 22-41 gave a weak positive reaction and fractions 42-57 gave a strong positive reaction to the anti-MiAMP2c antiserum. Fractions that showed antifungal activity against *S.sclerotiorum* at 50 µg/ml and 10 µg/ml are indicated by arrows on the chromatogram.

Using the approach above, several active fractions (termed SsAMP1 and SsAMP2) were obtained which were assessed for their antifungal activity against *Sclerotinia sclerotiorum, Alternaria brassicola, Leptosphaeria maculans, Verticilium dahliae* and *Fusarium oxysporum*. Bioassays were carried out as described in Example 2 and results shown in Example 15. Another fragment which reacted with MiAMP2 antiserum was purified and sequenced (SsAMP3) but insufficient protein was available for characterisation of antimicrobial activity. Partial sequences obtained from these proteins are shown in Figure 4 (SEQ ID NOS: 26, 27 and 28). Full sequencing of the peptides or cloning of cDNAs encoding the seed storage proteins from this species will reveal the extent of homology between these peptides and MiAMP2-series homologues.

Example 14

Synthesis of small fragments of MiAMP2c

In an effort to determine if the full MiAMP2c molecule was absolutely necessary for the protein to exhibit antimicrobial activity, two separate peptides were chemically synthesized by Auspep Pty. Ltd. (Australia). For each peptide, the cysteine residues were changed to alanine residues so that disulfide bonds were no longer capable of being formed between two separate protein chains. Tyrosine residues were also changed to alanine since it was expected that tyrosine also participated in the helix-turn-helix stabilization and this would not be needed in the synthetic peptides lacking one of the helices. Alanine is also favorable to the formation of alpha-helices so it should not interfere with the native helical structure to a large degree. Peptide one is comprised of 22 amino acids from 118 to 139 in the amino acid sequence of clone 3 (sequence: RQRDP QQQAE QAQKR AQRRE TE, SEQUENCE ID NO: 9). Peptide 2 is 25 amino acids in length and runs from 140 to 164 in clone 3 (sequence: PRHMQ IAQQR AERRA EKEKR KQQKR, SEQ ID NO: 10). Peptides 1 and 2 are labeled MiAMP2c pep1 and MiAMP2c pep2 respectively. These peptides were resuspended in Milli-Q water and bioassayed against a number of fungi. As seen in Table 2, peptide 2 has inhibitory activity against a variety of fungi whereas peptide 1 exhibited little or no activity. Mixtures of peptide 1 and peptide 2 exhibit similar levels of activity as seen with peptide 2 alone indicating that only peptide 2 is exhibiting activity. The fact that peptide 2 exhibits antimicrobial activity in the absence of the helix-turn-helix structure exhibited by MiAMP2c reveals that the helix-turn-helix structure is not absolutely necessary for the peptides to retain activity. Nevertheless,

peptide 2 did not exhibit the same degree of activity on a molar basis as MiAMP2c (whole fragment) indicating that the helix-turn-helix structure is important for maximal expression of antimicrobial activity by the fragments involved. It is also expected that the helix-turn-helix structure will confer greater stability to the MiAMP2 homologues, thus rendering these proteins less susceptible to proteolytic cleavage and other forms of degredation. Greater stability would lead to maintaining antimicrobial activity over a longer period of time.

Example 15

Antifungal activity of MiAMP2c homologues and fragment(s)

MiAMP2c and each of the various MiAMP2 homologues were tested against a variety of fungi as concentrations ranging from 2 to 50 μg/ml. Table 1 shows the IC_{50} value of pure MiAMP2c against various fungi and bacteria. In the table, the ">50" indicates that 50% inhibition of the fungus was not achieved at 50 μg/ml which was the highest concentration tested. The abbreviation "ND" indicates that the test was not performed or that results could not be interpreted. The antimicrobial activity of MiAMP2c was also tested in the presence of 1 mM Ca^{2+} in the test medium and the IC_{50} values for these tests are given in the right-hand column. As can be seen in the table, the inhibitory activity of MiAMP2c is greatly reduced (although not eliminated) in the presence of Ca^{2+}.

Table 1

Concentrations of MiAMP2c at which 50% inhibition of growth was observed

Organism	IC_{50} (μg/ml)	IC_{50} + Ca^{2+} (μg/ml)
Alternaria helianthi	5-10	ND
Candida albicans	>50	>50
Ceratocystis paradoxa	20-50	>50
Cercospora nicotianae	5-10	5-10
Clavibacter michiganensis	50	>50
Chalara elegans	2-5	10-20
Fusarium oxysporum	10	20-50
Sclerotinia sclerotiorum	20-50	>50
Phytophthora cryptogea	5-10	10-25
Phytophthora parasitica nicotiana	10-20	>50

Verticillium dahliae	5-10	>50
Ralstonia solanacearum	>50	>50
Pseudomonas syringae tabaci	>50	>50
Saccharomyces cerevisiae	20-50	>50
Escherichia coli	>50	>50

Table 2 shows the the antimicrobial activity of various homologues and fragments of MiAMP2c. In the table, the following abbreviations are used: Ab, *Alternaria brassicola*; Cp: *Ceratocystis paradoxa*; Foc: *Fusarium oxysporum*; Lm: *Leptosphaeria maculans*; Ss: *Sclerotinia sclerotiorum*; Vd: *Verticillium dahliae*. The ">50" indicates that concentrations higher than 50 µg/ml were not tested so that an IC_{50} value could not be established. A blank space indicates that the test was not performed or that results could not be interpreted.

The TcAMP1 and 2 used for the results presented in Table 2 were derived from cocoa vicilin (Examples 10 and 11). SsAMP1 and 2 show reactivity with MiAMP2c antibodies and also exhibit antimicrobial activity as seen in the table below. The versions of MiAMP2a, b and d as well as TcAMP1 and TcAMP2 tested in the bioassays all contain a His tag fusion resulting from expression in the vector pET16b. MiAMP2c pep1 and 2 are the N and C terminal regions, respectively, of MiAMP2c antimicrobial peptide as specified in Example 14 above. The concentration value listed for 'MiAMP2c pep1+2' is the concentration of each individual peptide in the mixture. It should be remembered that MiAMP2c pep1 and pep2 are both about ½ the size of MiAMP2c; comparisons of the activity of these peptides with the MiAMP2c protein should, therefore, be made on a molar basis rather than on a strict µg/ml concentration basis. Peptides were only tested in media A which did not contain added Ca^{2+}.

Table 2

IC_{50} values (µg/ml) of MiAMP2 related proteins against various fungi

Peptide tested	Fungus used in bioassy					
	Ab	Cp	Foc	Lm	Ss	Vd
MiAMP2a			5-10	2.5-5	5-10	
MiAMP2b			2.5	2.5	5-10	
MiAMP2c		20-50	10		20-50	5-10
MiAMP2d			5	2.5	5-10	
MiAMP2c pep1			100		>50	

MiAMP2c pep2			10-20	10-20	50	10-20
MiAMP2c pep1+2			10-25		50	
TcAMP1		10	5-10	2-5	10	5-20
TcAMP2		5-10	5-10	2-5	5	5-20
SsAMP1			20-50	20-50	20-50	10-20
SsAMP2	20-50		>50	>50	>50	>50

It is worthy of note that while the TcAMP1 and 2 sequences are readily available in the public data bases, no antimicrobial activity had ever been assigned to them. These sequences were derived from much larger proteins involved in seed storage functions. The inventors have thus described a completely new activity for a small portion of the overall cocoa vicilin molecules. The activity of cotton fragments 1, 2, and 3 has been exemplified by other authors (Chung, R. P.T. *et al.* [1997] *Plant Science* 127:1-16).

Example 16

Construction of the plant transfomation vector PCV91-MiAMP2c

The expression vector pPCV91-MiAMP2c (Figure 14) contains the full coding region of the MiAMP2c (Example 7) DNA flanked at it 5' end by the strong constitutive promoter of 35S RNA from the cauliflower mosaic virus (pCaMV35S) (Odel *et al.*, [1985] *Nature* 313: 810-812) with a quadruple-repeat enhancer element (e-35S) to allow for high transcriptional activity (Kay *et al.* [1987] *Science* 236:1299-1302). The coding region of MiAMP2c DNA is flanked at its 3' end by the polyadenylation sequence of 35S RNA of the cauliflower mosaic virus (pA35S). The plasmid backbone of this vector is the plasmid pPCV91 (Walden, R. *et al.* [1990] *Methods Mol. Cell. Biol.* 1:175-194). The plasmid also contains other elements useful for plant transformation such as an ampicillin resistance gene (bla) and a hygromycin resistance gene (hph) driven by the nos promoter (pnos). These and other features allow for selection in various cloning and transformation procedures. The plasmid pPCV91-MiAMP2c was constructed as follows: A cloned fragment encoding MiAMP2c (Example 7) was digested using restriction enzymes to release the MiAMP2c gene fragment containing a synthetic leader sequence.. The binary vector pPCV91 was digested with the restriction enzyme *Bam* HI. Both the MiAMP2c DNA fragment containing and the binary vector were ligated using T4 DNA ligase to produce pPCV91-MiAMP2c binary vector for plant transformation (Figure 12).

Using this approach, other homologues of MiAMP2c can be expressed in plants. Not only can individual homologues be expressed, but they may be expressed in combination with other proteins as fusion proteins or as portions of larger precursor proteins. For example, it is possible to express

the N-terminal region of MiAMP2 clone 1 (amino acids 1 to ~246) which contains a signal peptide and the hydrophilic region containing four antimicrobial segments. Transgenic plants can then be assessed to examine whether the individual fragments are being processed into the expected fragments by the processing machinery already present in the plant cells. It is also possible to express the entire MiAMP2 clone 1 (amino acids 1 to 666) and to examine the processing of the entire protein when expressed in transgenic plants. Homologous regions from other sequences can also be used in multiple combinations with, for example, ten (10) or more MiAMP2-like fragments expressed as one large fusion protein with acidic cleavage sites located as proper locations between each of the fragments. As well as linking MiAMP2 fragments together, it would also be possible to link MiAMP2 fragments to other useful proteins for expression in plants.

Example 17

Transgenic plants expressing MiAMP2c (or related fragments)

The disarmed *Agrobacterium tumefaciens* strain GV3101 (pMP90RK) (Koncz, Cs.[1986] *Mol. Gen. Genet.* 204:383-396) was transformed with the vector pPCV91-MiAMP2c (Example 16) using the method of Walkerpeach *et al.* (Plant Mol. Biol. Manual B1:1-19 [1994]) adapted from Van Haute *et al* (*EMBO J.* 2:411-417 1983]).

Tobacco transformation was carried out using leaf discs of *Nicotiana tabacum* based on the method of Horsch *et al.* (*Science* 227:1229-1231 [1985]) and co-culturing strains containing pPCV91-MiAMP2c. After co-cultivation of *Agrobacterium* and tobacco leaf disks, transgenic plants (transformed with pPCV91-MiAMP2c) were regenerated on media containing 50 µg/ml hygromycin and 500 µg/ml Cefotaxime. These transgenic plants were analysed for expression of the newly-introduced genes using standard western blotting techniques (Figure 15). Figure 15 shows a western blot of extracts from trangenic tobacco carrying the construct for MiAMP2c from example 16. Lane 1 contains pure MiAMP2c as a standard, lanes 2 and 3 contain extracts from transgenic plants carrying the pPCV91-MiAMP2c construct. As can be see in the figure, faint bands are present at approximately the correct molecular weight, indicating that the transgenic plants appear to be expressing the MiAMP2c protein. Plants capable of constitutive expression of the introduced genes may be selected and self-pollinated to give seed. F1 seedlings of the transgenic plants may be further analysed.

Example 18

MiAMP2c Homologues

Every homologue of MiAMP2c that has been tested has exhibited some antimicrobial activity. This evidence indicates that other homologues will also exhibit antimicrobial activity. These homologues include fragments from 1) peanut (Burks, A.W. *et al.* [1995] *J. Clin. Invest.* 96 (4),

1715-1721), 2) maize (Belanger, F.C. and Kriz, A.L.[1991] *Genetics* 129 (3), 863-872), 3) barley (Heck, G.R. *et al.* [1993] *Mol. Gen. Genet.* 239 (1-2), 209-218), and 4) soybean (Sebastiani, F.L. *et al.* [1990] *Plant Mol. Biol.* 15 (1), 197-201). (see SEQ ID NOS: 21, 22, 24, and 25). Other sequences derived from seed storage proteins of the 7S class are also expected to yield homologues of MiAMP2 proteins.

SEQUENCE LISTINGS

(1) GENERAL INFORMATION:

(i)APPLICANT:
(A) NAME: COOPERATIVE RESEARCH CENTRE FOR TROPICAL PLANT PATHOLOGY
(B) STREET: The University of Queensland
(C) CITY: St Lucia
(D) STATE: Queensland
(E) COUNTRY: Australia
(F) POSTAL CODE (ZIP): 4067

(ii) TITLE OF INVENTION: Antimicrobial Protein

(iii) NUMBER OF SEQUENCES: 28

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 666 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Macadamia integrifolia
(F) TISSUE TYPE: Seeds

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

Met Ala Ile Asn Thr Ser Asn Leu Cys Ser Leu Leu Phe Leu Leu Ser
1 5 10 15

Leu Phe Leu Leu Ser Thr Thr Val Ser Leu Ala Glu Ser Glu Phe Asp
20 25 30

Arg Gln Glu Tyr Glu Glu Cys Lys Arg Gln Cys Met Gln Leu Glu Thr
35 40 45

Ser Gly Gln Met Arg Arg Cys Val Ser Gln Cys Asp Lys Arg Phe Glu
50 55 60

Glu Asp Ile Asp Trp Ser Lys Tyr Asp Asn Gln Glu Asp Pro Gln Thr

65 70 75 80

Glu Cys Gln Gln Cys Gln Arg Arg Cys Arg Gln Gln Glu Ser Gly Pro
85 90 95

Arg Gln Gln Gln Tyr Cys Gln Arg Arg Cys Lys Glu Ile Cys Glu Glu
100 105 110

Glu Glu Glu Tyr Asn Arg Gln Arg Asp Pro Gln Gln Gln Tyr Glu Gln
115 120 125

Cys Gln Lys His Cys Gln Arg Arg Glu Thr Glu Pro Arg His Met Gln
130 135 140

Thr Cys Gln Gln Arg Cys Glu Arg Arg Tyr Glu Lys Glu Lys Arg Lys
145 150 155 160

Gln Gln Lys Arg Tyr Glu Glu Gln Gln Arg Glu Asp Glu Glu Lys Tyr
165 170 175

Glu Glu Arg Met Lys Glu Glu Asp Asn Lys Arg Asp Pro Gln Gln Arg
180 185 190

Glu Tyr Glu Asp Cys Arg Arg Arg Cys Glu Gln Gln Glu Pro Arg Gln
195 200 205

Gln His Gln Cys Gln Leu Arg Cys Arg Glu Gln Gln Arg Gln His Gly
210 215 220

Arg Gly Gly Asp Met Met Asn Pro Gln Arg Gly Gly Ser Gly Arg Tyr
225 230 235 240

Glu Glu Gly Glu Glu Glu Gln Ser Asp Asn Pro Tyr Tyr Phe Asp Glu
245 250 255

Arg Ser Leu Ser Thr Arg Phe Arg Thr Glu Glu Gly His Ile Ser Val
260 265 270

Leu Glu Asn Phe Tyr Gly Arg Ser Lys Leu Leu Arg Ala Leu Lys Asn
275 280 285

Tyr Arg Leu Val Leu Leu Glu Ala Asn Pro Asn Ala Phe Val Leu Pro
290 295 300

Thr His Leu Asp Ala Asp Ala Ile Leu Leu Val Ile Gly Gly Arg Gly
305 310 315 320

Ala Leu Lys Met Ile His His Asp Asn Arg Glu Ser Tyr Asn Leu Glu
325 330 335

Cys Gly Asp Val Ile Arg Ile Pro Ala Gly Thr Thr Phe Tyr Leu Ile
340 345 350

Asn Arg Asp Asn Asn Glu Arg Leu His Ile Ala Lys Phe Leu Gln Thr
355 360 365

Ile Ser Thr Pro Gly Gln Tyr Lys Glu Phe Phe Pro Ala Gly Gly Gln
370 375 380

Asn Pro Glu Pro Tyr Leu Ser Thr Phe Ser Lys Glu Ile Leu Glu Ala
385 390 395 400

Ala Leu Asn Thr Gln Thr Glu Lys Leu Arg Gly Val Phe Gly Gln Gln
405 410 415

Arg Glu Gly Val Ile Ile Arg Ala Ser Gln Glu Gln Ile Arg Glu Leu
420 425 430

Thr Arg Asp Asp Ser Glu Ser Arg His Trp His Ile Arg Arg Gly Gly
435 440 445

Glu Ser Ser Arg Gly Pro Tyr Asn Leu Phe Asn Lys Arg Pro Leu Tyr
450 455 460

Ser Asn Lys Tyr Gly Gln Ala Tyr Glu Val Lys Pro Glu Asp Tyr Arg
465 470 475 480

Gln Leu Gln Asp Met Asp Leu Ser Val Phe Ile Ala Asn Val Thr Gln
485 490 495

Gly Ser Met Met Gly Pro Phe Phe Asn Thr Arg Ser Thr Lys Val Val
500 505 510

Val Val Ala Ser Gly Glu Ala Asp Val Glu Met Ala Cys Pro His Leu
515 520 525

Ser Gly Arg His Gly Gly Arg Gly Gly Gly Lys Arg His Glu Glu Glu
530 535 540

Glu Asp Val His Tyr Glu Gln Val Arg Ala Arg Leu Ser Lys Arg Glu
545 550 555 560

Ala Ile Val Val Leu Ala Gly His Pro Val Val Phe Val Ser Ser Gly
565 570 575

Asn Glu Asn Leu Leu Leu Phe Ala Phe Gly Ile Asn Ala Gln Asn Asn
580 585 590

His Glu Asn Phe Leu Ala Gly Arg Glu Arg Asn Val Leu Gln Gln Ile
595 600 605

Glu Pro Gln Ala Met Glu Leu Ala Phe Ala Ala Pro Arg Lys Glu Val
610 615 620

Glu Glu Ser Phe Asn Ser Gln Asp Gln Ser Ile Phe Phe Pro Gly Pro
625 630 635 640

Arg Gln His Gln Gln Gln Ser Pro Arg Ser Thr Lys Gln Gln Gln Pro
645 650 655

Leu Val Ser Ile Leu Asp Phe Val Gly Phe
660 665

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2171 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Macadamia integrifolia
(F) TISSUE TYPE: Seeds

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION:1..85

(x) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:86..1999

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

ATGGCGATCA ATACATCAAA TTTATGTTCT CTTCTCTTTC TCCTTTCACT CTTCCTTCTG 60

TCTACGACAG TGTCTCTTGC TGAAAGTGAA TTTGACAGGC AGGAATATGA GGAGTGCAAA 120

CGGCAATGCA TGCAGTTGGA GACATCAGGC CAGATGCGTC GGTGTGTGAG TCAGTGCGAT 180

AAGAGATTTG AAGAGGATAT AGATTGGTCT AAGTATGATA ACCAAGAGGA TCCTCAGACG 240

GAATGCCAAC AATGCCAGAG GCGATGCAGG CAGCAGGAGA GTGGCCCACG TCAGCAACAA 300

TACTGCCAAC GACGCTGCAA GGAAATATGT GAAGAAGAAG AAGAATATAA CCGACAACGT 360

GATCCACAGC AGCAATACGA GCAATGTCAG AAGCACTGCC AACGGCGCGA GACAGAGCCA 420

CGTCACATGC AAACATGTCA ACAACGCTGC GAGAGGAGAT ATGAAAAGGA GAAACGTAAG 480

CAACAAAAGA GATATGAAGA GCAACAACGT GAAGACGAAG AGAAATATGA AGAGCGAATG 540

AAGGAAGAAG ATAACAAACG CGATCCACAA CAAAGAGAGT ACGAAGACTG CCGGAGGCGC 600

TGCGAACAAC AGGAGCCACG TCAGCAGCAC CAGTGCCAGC TAAGATGCCG AGAGCAGCAG 660

AGGCAACACG GCCGAGGTGG CGATATGATG AACCCTCAGA GGGGAGGCAG CGGCAGATAC 720

GAGGAGGGAG AAGAGGAGCA AAGCGACAAC CCCTACTACT TCGACGAACG AAGCTTAAGT 780

ACAAGGTTCA GGACCGAGGA AGGCCACATC TCAGTTCTGG AGAACTTCTA TGGTAGATCC 840

AAGCTTCTAC GCGCACTAAA AAACTATCGC TTGGTGCTCC TCGAGGCTAA CCCCAACGCC 900

TTCGTGCTCC CTACCCACTT GGATGCAGAT GCCATTCTCT TGGTCATAGG AGGGAGAGGA 960

GCCCTCAAAA TGATCCACCA CGACAACAGA GAATCCTACA ACCTCGAGTG TGGAGACGTA 1020

ATCAGAATCC CAGCTGGAAC CACATTCTAC TTAATCAACC GAGACAACAA CGAGAGGCTC 1080

CACATAGCCA AGTTCTTACA GACCATATCC ACTCCTGGCC AATACAAGGA ATTCTTCCCA 1140

GCTGGAGGCC AAAACCCAGA GCCGTACCTC AGTACCTTCA GCAAAGAGAT TCTCGAGGCT 1200

GCGCTCAACA CACAAACAGA GAAGCTGCGT GGGGTGTTTG GACAGCAAAG GGAGGGAGTG 1260

ATAATTAGGG CGTCACAGGA GCAGATCAGG GAGTTGACTC GAGATGACTC AGAGTCACGA 1320

CACTGGCATA TAAGGAGAGG TGGTGAATCA AGCAGGGGAC CTTACAATCT GTTCAACAAA 1380

AGGCCACTGT ACTCCAACAA ATACGGTCAA GCCTACGAAG TCAAACCTGA GGACTACAGG 1440

CAACTCCAAG ACATGGACTT ATCGGTTTTC ATAGCCAACG TCACCCAGGG ATCCATGATG 1500

GGTCCCTTCT TCAACACTAG GTCTACAAAG GTGGTAGTGG TGGCTAGTGG AGAGGCAGAT 1560

GTGGAAATGG CATGCCCTCA CTTGTCGGGA AGACACGGCG GCCGCGGTGG AGGAAAAAGG 1620

CATGAGGAGG AAGAGGATGT GCACTATGAG CAGGTTAGAG CACGTTTGTC GAAGAGAGAG 1680

GCCATTGTTG TTCTGGCAGG TCATCCCGTC GTCTTCGTTT CATCCGGAAA CGAGAACCTG 1740

CTGCTTTTTG CATTTGGAAT CAATGCCCAA AACAACCACG AGAACTTCCT CGCGGGGAGA 1800

GAGAGGAACG TGCTGCAGCA GATAGAGCCA CAGGCAATGG AGCTAGCGTT TGCCGCTCCA 1860

AGGAAAGAGG TAGAAGAGTC ATTTAACAGC CAGGACCAGT CTATCTTCTT TCCTGGGCCC 1920

AGGCAGCACC AGCAACAGTC GCCCCGCTCC ACCAAGCAAC AACAGCCTCT CGTCTCCATT 1980

CTGGACTTCG TTGGCTTCTA AAGTTCCACA AAAAAGAGTG TGTTATGTAG TATAGGTTAG 2040

TAGCTCCTAG CTCGGTGTAT GAGAGTGGTA AGAGACTAAG ACGCTAAATC CCTAAGTAAC 2100

TAACCTGGCG AGCTTGCGTG TATGCAAATA AAGAGGAACA GCTTTCCAAC TTTAAAAAAA 2160

AAAAAAAAAA A 2171

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 666 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Macadamia integrifolia
(F) TISSUE TYPE: Seeds

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION:1..28

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:29..666

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
Met Ala Ile Asn Thr Ser Asn Leu Cys Ser Leu Leu Phe Leu Leu Ser
1               5                  10                  15

Leu Phe Leu Leu Ser Thr Thr Val Ser Leu Ala Glu Ser Glu Phe Asp
            20                  25                  30

Arg Gln Glu Tyr Glu Glu Cys Lys Arg Gln Cys Met Gln Leu Glu Thr
        35                  40                  45

Ser Gly Gln Met Arg Arg Cys Val Ser Gln Cys Asp Lys Arg Phe Glu
    50                  55                  60

Glu Asp Ile Asp Trp Ser Lys Tyr Asp Asn Gln Asp Asp Pro Gln Thr
65                  70                  75                  80

Asp Cys Gln Gln Cys Gln Arg Arg Cys Arg Gln Gln Glu Ser Gly Pro
                85                  90                  95

Arg Gln Gln Gln Tyr Cys Gln Arg Arg Cys Lys Glu Ile Cys Glu Glu
            100                 105                 110

Glu Glu Glu Tyr Asn Arg Gln Arg Asp Pro Gln Gln Gln Tyr Glu Gln
        115                 120                 125

Cys Gln Glu Arg Cys Gln Arg His Glu Thr Glu Pro Arg His Met Gln
    130                 135                 140

Thr Cys Gln Gln Arg Cys Glu Arg Arg Tyr Glu Lys Glu Lys Arg Lys
145                 150                 155                 160

Gln Gln Lys Arg Tyr Glu Glu Gln Gln Arg Glu Asp Glu Glu Lys Tyr
                165                 170                 175

Glu Glu Arg Met Lys Glu Glu Asp Asn Lys Arg Asp Pro Gln Gln Arg
            180                 185                 190

Glu Tyr Glu Asp Cys Arg Arg Arg Cys Glu Gln Gln Glu Pro Arg Gln
        195                 200                 205

Gln Tyr Gln Cys Gln Arg Arg Cys Arg Glu Gln Gln Arg Gln His Gly
    210                 215                 220
```

Arg Gly Gly Asp Leu Ile Asn Pro Gln Arg Gly Gly Ser Gly Arg Tyr
225 230 235 240

Glu Glu Gly Glu Glu Lys Gln Ser Asp Asn Pro Tyr Tyr Phe Asp Glu
245 250 255

Arg Ser Leu Ser Thr Arg Phe Arg Thr Glu Glu Gly His Ile Ser Val
260 265 270

Leu Glu Asn Phe Tyr Gly Arg Ser Lys Leu Leu Arg Ala Leu Lys Asn
275 280 285

Tyr Arg Leu Val Leu Leu Glu Ala Asn Pro Asn Ala Phe Val Leu Pro
290 295 300

Thr His Leu Asp Ala Asp Ala Ile Leu Leu Val Thr Gly Gly Arg Gly
305 310 315 320

Ala Leu Lys Met Ile His Arg Asp Asn Arg Glu Ser Tyr Asn Leu Glu
325 330 335

Cys Gly Asp Val Ile Arg Ile Pro Ala Gly Thr Thr Phe Tyr Leu Ile
340 345 350

Asn Arg Asp Asn Asn Glu Arg Leu His Ile Ala Lys Phe Leu Gln Thr
355 360 365

Ile Ser Thr Pro Gly Gln Tyr Lys Glu Phe Phe Pro Ala Gly Gly Gln
370 375 380

Asn Pro Glu Pro Tyr Leu Ser Thr Phe Ser Lys Glu Ile Leu Glu Ala
385 390 395 400

Ala Leu Asn Thr Gln Ala Glu Arg Leu Arg Gly Val Leu Gly Gln Gln
405 410 415

Arg Glu Gly Val Ile Ile Ser Ala Ser Gln Glu Gln Ile Arg Glu Leu
420 425 430

Thr Arg Asp Asp Ser Glu Ser Arg Arg Trp His Ile Arg Arg Gly Gly
435 440 445

Glu Ser Ser Arg Gly Pro Tyr Asn Leu Phe Asn Lys Arg Pro Leu Tyr
450 455 460

Ser Asn Lys Tyr Gly Gln Ala Tyr Glu Val Lys Pro Glu Asp Tyr Arg
465 470 475 480

Gln Leu Gln Asp Met Asp Val Ser Val Phe Ile Ala Asn Ile Thr Gln
485 490 495

Gly Ser Met Met Gly Pro Phe Phe Asn Thr Arg Ser Thr Lys Val Val
500 505 510

Val Val Ala Ser Gly Glu Ala Asp Val Glu Met Ala Cys Pro His Leu

515 520 525

Ser Gly Arg His Gly Gly Arg Arg Gly Gly Lys Arg His Glu Glu Glu
530 535 540

Glu Asp Val His Tyr Glu Gln Val Lys Ala Arg Leu Ser Lys Arg Glu
545 550 555 560

Ala Ile Val Val Pro Val Gly His Pro Val Val Phe Val Ser Ser Gly
565 570 575

Asn Glu Asn Leu Leu Leu Phe Ala Phe Gly Ile Asn Ala Gln Asn Asn
580 585 590

His Glu Asn Phe Leu Ala Gly Arg Glu Arg Asn Val Leu Gln Gln Ile
595 600 605

Glu Pro Gln Ala Met Glu Leu Ala Phe Ala Ala Pro Arg Lys Glu Val
610 615 620

Glu Glu Leu Phe Asn Ser Gln Asp Glu Ser Ile Phe Phe Pro Gly Pro
625 630 635 640

Arg Gln His Gln Gln Gln Ser Ser Arg Ser Thr Lys Gln Gln Gln Pro
645 650 655

Leu Val Ser Ile Leu Asp Phe Val Gly Phe
660 665

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2171 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Macadamia integrifolia
(F) TISSUE TYPE: Seeds

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION:1..86

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:87..1999

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

ATGGCGATCA ATACATCAAA TTTATGTTCT CTTCTCTTTC TCCTTTCCCT CTTCCTTCTG 60

TCAACGACAG TGTCTCTTGC TGAAAGTGAA TTTGACAGGC AGGAATATGA GGAGTGCAAA 120

CGGCAATGCA TGCAGTTGGA GACATCAGGC CAGATGCGTC GGTGTGTGAG TCAGTGCGAT 180

AAGAGATTTG AAGAGGATAT AGATTGGTCT AAGTATGATA ACCAAGACGA TCCTCAGACG 240

GATTGCCAAC AATGCCAGAG GCGATGCAGG CAGCAGGAGA GTGGCCCACG TCAGCAACAA 300

TACTGCCAAC GACGCTGCAA GGAAATATGT GAAGAAGAAG AAGAATATAA CCGACAACGT 360

GATCCACAGC AGCAATACGA GCAATGTCAG GAGCGCTGCC AACGGCACGA GACAGAGCCA 420

CGTCACATGC AAACATGTCA ACAACGCTGC GAGAGGAGAT ATGAAAAGGA GAAACGTAAG 480

CAACAAAAGA GATATGAAGA GCAACAACGT GAAGACGAAG AGAAATATGA AGAGCGAATG 540

AAGGAAGAAG ATAACAAACG CGATCCACAA CAAAGAGAGT ACGAAGACTG CCGGAGGCGC 600

TGCGAACAAC AGGAGCCACG TCAGCAGTAC CAGTGCCAGC GAAGATGCCG AGAGCAGCAG 660

AGGCAACACG GCCGAGGTGG TGATTTGATT AACCCTCAGA GGGGAGGCAG CGGCAGATAC 720

GAGGAGGGAG AAGAGAAGCA AAGCGACAAC CCCTACTACT TCGACGAACG AAGCTTAAGT 780

ACAAGGTTCA GGACCGAGGA AGGCCACATC TCAGTTCTGG AGAACTTCTA TGGTAGATCC 840

AAGCTTCTAC GCGCACTAAA AAACTATCGC TTGGTGCTCC TCGAGGCTAA CCCCAACGCC 900

TTCGTGCTCC CTACCCACTT GGACGCAGAT GCCATTCTCT TGGTCACCGG AGGGAGAGGA 960

GCCCTCAAAA TGATCCACCG TGACAACAGA GAATCCTACA ACCTCGAGTG TGGAGACGTA 1020

ATCAGAATCC CAGCTGGAAC CACATTCTAC TTAATCAACC GAGACAACAA CGAGAGGCTC 1080

CACATAGCCA AGTTCTTACA GACCATATCC ACTCCTGGCC AATACAAGGA ATTCTTCCCA 1140

GCTGGAGGCC AAAACCCAGA GCCGTACCTC AGTACCTTCA GCAAAGAGAT TCTCGAGGCT 1200

GCGCTCAACA CACAAGCAGA GAGGCTGCGT GGGGTGCTTG GACAGCAAAG GGAGGGAGTG 1260

ATAATTAGTG CGTCACAGGA GCAGATCAGG GAGTTGACTC GAGATGACTC AGAGTCACGA 1320

CGCTGGCATA TAAGGAGAGG TGGTGAATCA AGCAGGGGAC CTTACAATCT GTTCAACAAA 1380

AGGCCACTGT ACTCCAACAA ATACGGTCAA GCCTACGAAG TCAAACCTGA GGACTACAGG 1440

CAACTCCAAG ACATGGACGT ATCGGTTTTC ATAGCCAACA TCACCCAGGG ATCCATGATG 1500

GGTCCCTTCT TCAACACTAG GTCTACAAAG GTGGTAGTGG TGGCTAGTGG AGAGGCAGAT 1560

GTGGAAATGG CATGCCCTCA CTTGTCGGGA AGACACGGCG GCCGCCGTGG AGGGAAAAGG 1620

CATGAGGAGG AAGAGGATGT GCACTATGAG CAGGTTAAAG CACGTTTGTC GAAGAGAGAG 1680

GCCATTGTTG TTCCGGTAGG TCATCCCGTC GTCTTCGTTT CATCCGGAAA CGAGAACCTG 1740

CTGCTTTTTG CATTTGGAAT CAATGCCCAA AACAACCACG AGAACTTCCT CGCGGGGAGA 1800

GAGAGGAACG TGCTGCAGCA GATAGAGCCA CAGGCAATGG AGCTAGCGTT TGCCGCTCCA 1860

AGGAAAGAGG TAGAAGAGTT ATTTAACAGC CAGGACGAGT CTATCTTCTT TCCTGGGCCC 1920

AGGCAGCACC AGCAACAGTC TTCCCGCTCC ACCAAGCAAC AACAGCCTCT CGTCTCCATT 1980

CTGGACTTCG TTGGCTTCTA AAGTTCTACA AAAAAGAGTG TGTTATGTAG TATAGGTTAG 2040

TAGCTCCTAG CTCGGTGTAT GCGAGTGGTA AGAGACCAAG ACGCTAAATC CCTAAGTAAC 2100

TAACCTGGCG AGCTTGCGTG TATGCAAATA AAGAGGAACA GCTTTCCAAC TTTAAAAAAA 2160

AAAAAAAAAA A 2171

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 625 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Macadamia integrifolia
(F) TISSUE TYPE: Seeds

(ix) FEATURE:
(A) NAME/KEY: partial mat_peptide
(B) LOCATION:1..625

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Gln Cys Met Gln Leu Glu Thr Ser Gly Gln Met Arg Arg Cys Val Ser
1 5 10 15

Gln Cys Asp Lys Arg Phe Glu Glu Asp Ile Asp Trp Ser Lys Tyr Asp
20 25 30

Asn Gln Glu Asp Pro Gln Thr Glu Cys Gln Gln Cys Gln Arg Arg Cys
35 40 45

Arg Gln Gln Glu Ser Asp Pro Arg Gln Gln Gln Tyr Cys Gln Arg Arg
50 55 60

Cys Lys Glu Ile Cys Glu Glu Glu Glu Glu Tyr Asn Arg Gln Arg Asp
65 70 75 80

Pro Gln Gln Gln Tyr Glu Gln Cys Gln Lys Arg Cys Gln Arg Arg Glu
85 90 95

Thr Glu Pro Arg His Met Gln Ile Cys Gln Gln Arg Cys Glu Arg Arg
100 105 110

Tyr Glu Lys Glu Lys Arg Lys Gln Gln Lys Arg Tyr Glu Glu Gln Gln
115 120 125

Arg Glu Asp Glu Glu Lys Tyr Glu Glu Arg Met Lys Glu Gly Asp Asn
130 135 140

Lys Arg Asp Pro Gln Gln Arg Glu Tyr Glu Asp Cys Arg Arg His Cys
145 150 155 160

Glu Gln Gln Glu Pro Arg Leu Gln Tyr Gln Cys Gln Arg Arg Cys Gln
165 170 175 180

Glu Gln Gln Arg Gln His Gly Arg Gly Gly Asp Leu Met Asn Pro Gln
185 190 195

Arg Gly Gly Ser Gly Arg Tyr Glu Glu Gly Glu Glu Lys Gln Ser Asp
200 205 210

Asn Pro Tyr Tyr Phe Asp Glu Arg Ser Leu Ser Thr Arg Phe Arg Thr
215 220 225

Glu Glu Gly His Ile Ser Val Leu Glu Asn Phe Tyr Gly Arg Ser Lys
230 235 240 245

Leu Leu Arg Ala Leu Lys Asn Tyr Arg Leu Val Leu Leu Glu Ala Asn
250 255 260

Pro Asn Ala Phe Val Leu Pro Thr His Leu Asp Ala Asp Ala Ile Leu
265 270 275

Leu Val Ile Gly Gly Arg Gly Ala Leu Lys Met Ile His Arg Asp Asn
280 285 290

Arg Glu Ser Tyr Asn Leu Glu Cys Gly Asp Val Ile Arg Ile Pro Ala
295 300 305

Gly Thr Thr Phe Tyr Leu Ile Asn Arg Asp Asn Asn Glu Arg Leu His
310 315 320 325

Ile Ala Lys Phe Leu Gln Thr Ile Ser Thr Pro Gly Gln Tyr Lys Glu
330 335 340

Phe Phe Pro Ala Gly Gly Gln Asn Pro Glu Pro Tyr Leu Ser Thr Phe
345 350 355

Ser Lys Glu Ile Leu Glu Ala Ala Leu Asn Thr Gln Thr Glu Arg Leu
360 365 370

Arg Gly Val Leu Gly Gln Gln Arg Glu Gly Val Ile Ile Arg Ala Ser
375 380 385

Gln Glu Gln Ile Arg Glu Leu Thr Arg Asp Asp Ser Glu Ser Arg Arg
390 395 400 405

Trp His Ile Arg Arg Gly Gly Glu Ser Ser Arg Gly Pro Tyr Asn Leu
410 415 420

Phe Asn Lys Arg Pro Leu Tyr Ser Asn Lys Tyr Gly Gln Ala Tyr Glu
425 430 435

Val Lys Pro Glu Asp Tyr Arg Gln Leu Gln Asp Met Asp Val Ser Val
440 445 450

Phe Ile Ala Asn Ile Thr Gln Gly Ser Met Met Gly Pro Phe Phe Asn
455 460 470

Thr Arg Ser Thr Lys Val Val Val Val Ala Ser Gly Glu Ala Asp Val
480 485 490 500

Glu Met Ala Cys Pro His Leu Ser Gly Arg His Gly Gly Arg Gly Gly
505 510 515

Gly Lys Arg His Glu Glu Glu Glu Glu Val His Tyr Glu Gln Val Arg
520 525 530

Ala Arg Leu Ser Lys Arg Glu Ala Ile Val Val Leu Ala Gly His Pro
535 540 545

Val Val Phe Val Ser Ser Gly Asn Glu Asn Leu Leu Leu Phe Ala Phe
550 555 560

Gly Ile Asn Ala Gln Asn Asn His Glu Asn Phe Leu Ala Gly Arg Glu
565 570 575 580

Arg Asn Val Leu Gln Gln Ile Glu Pro Gln Ala Met Glu Leu Ala Phe
585 590 595

Ala Ala Ser Arg Lys Glu Val Glu Glu Leu Phe Asn Ser Gln Asp Glu
600 605 610

Ser Ile Phe Phe Pro Gly Pro Arg Gln His Gln Gln Gln Ser Pro Arg
615 620 625

Ser Thr Lys Gln Gln Gln Pro Leu Val Ser Ile Leu Asp Phe Val Gly
630 635 640
Phe

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2140 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Macadamia integrifolia

(F) TISSUE TYPE: Seeds

(x) FEATURE:
(A) NAME/KEY: partial mat_peptide
(B) LOCATION:1..1875

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CAATGCATGC AGTTAGAGAC ATCAGGCCAG ATGCGTCGGT GTGTGAGTCA GTGCGATAAG 60

AGATTTGAAG AGGATATAGA TTGGTCTAAG TATGATAACC AAGAGGATCC TCAGACGGAA 120

TGCCAACAAT GCCAGAGGCG ATGCAGGCAG CAGGAGAGTG ACCCACGTCA GCAACAATAC 180

TGCCAACGAC GCTGCAAGGA AATATGTGAA GAAGAAGAAG AATATAACCG ACAACGTGAT 240

CCACAGCAGC AATACGAGCA ATGTCAGAAG CGCTGCCAAC GGCGCGAGAC AGAGCCACGT 300

CACATGCAAA TATGTCAACA ACGCTGCGAG AGGAGATATG AAAAGGAGAA ACGTAAGCAA 360

CAAAAGAGAT ATGAAGAGCA ACAACGTGAA GACGAAGAGA AATATGAAGA GCGAATGAAG 420

GAAGGAGATA ACAAACGCGA TCCACAACAA AGAGAGTACG AAGACTGCCG GCGGCACTGC 480

GAACAACAGG AGCCACGTCT GCAGTACCAG TGCCAGCGAA GATGCCAAGA GCAGCAGAGG 540

CAACACGGCC GAGGTGGCGA TTTGATGAAC CCTCAGAGGG GAGGCAGCGG CAGATACGAG 600

GAGGGAGAAG AGAAGCAAAG CGACAACCCC TACTACTTCG ACGAACGAAG CTTAAGTACA 660

AGGTTCAGGA CCGAGGAAGG CCACATCTCA GTTCTGGAGA ACTTCTATGG TAGATCCAAG 720

CTTCTACGCG CACTAAAAAA CTATCGCTTG GTGCTCCTCG AGGCTAACCC CAACGCCTTC 780

GTGCTCCCTA CCCACTTGGA TGCAGATGCC ATTCTCTTGG TCATCGGAGG GAGAGGAGCC 840

CTCAAAATGA TCCACCGTGA CAACAGAGAA TCCTACAACC TCGAGTGTGG AGACGTAATC 900

AGAATCCCAG CTGGAACCAC ATTCTACTTA ATCAACCGAG ACAACAACGA GAGGCTCCAC 960

ATAGCCAAGT TCTTACAGAC CATATCCACT CCTGGCCAAT ACAAGGAATT CTTCCCAGCT 1020

GGAGGCCAAA ACCCAGAGCC GTACCTCAGT ACCTTCAGCA AAGAGATTCT CGAGGCTGCG 1080

CTCAACACAC AAACAGAGAG GCTGCGTGGG GTGCTTGGAC AGCAAAGGGA GGGAGTGATA 1140

ATTAGGGCGT CACAGGAGCA GATCAGGGAG TTGACTCGAG ATGACTCAGA GTCACGACGC 1200

TGGCATATAA GGAGAGGTGG TGAATCAAGC AGGGGACCTT ACAATCTGTT CAACAAAAGG 1260

CCACTGTACT CCAACAAATA CGGTCAAGCC TACGAAGTCA AACCTGAGGA CTACAGGCAA 1320

CTCCAAGACA TGGACGTATC AGTTTTCATA GCCAACATCA CCCAGGGATC CATGATGGGT 1380

CCCTTCTTCA ACACTAGGTC TACAAAGGTG GTAGTGGTGG CTAGTGGAGA GGCAGATGTG 1440

GAAATGGCAT GCCCTCACTT GTCGGGAAGA CACGGCGGCC GCGGTGGAGG GAAAAGGCAT 1500

GAGGAGGAAG AGGAGGTGCA CTATGAGCAG GTTAGAGCAC GTTTGTCGAA GAGAGAGGCC 1560

ATTGTTGTTC TGGCAGGTCA TCCCGTCGTC TTCGTTTCAT CCGGAAACGA AAACCTGCTG 1620

CTTTTTGCAT TTGGAATCAA TGCCCAAAAC AACCACGAGA ACTTCCTCGC GGGGAGAGAG 1680

AGGAACGTGC TGCAGCAGAT AGAGCCACAG GCAATGGAGC TAGCGTTTGC CGCTTCAAGG 1740

AAAGAGGTAG AAGAGTTATT TAACAGCCAG GACGAGTCTA TCTTCTTTCC TGGGCCCAGG 1800

CAGCACCAGC AACAGTCGCC CCGCTCCACC AAGCAACAAC AGCCTCTCGT CTCCATTCTG 1860

GACTTCGTTG GCTTCTAAAG TTCTACAAAA AAGAGTGTGT TATGTAGTAT AGGTTAGTAG 1920

CTCCTAGCTC GGTGTATGAG AGTGGTAAGA GACTAAGACG CTAAATCCCT AAGTAACTAA 1980

CCTGGCGAGC TTGCGTGTAT GCAAATAAAG AGGAACAGCT TTCCAACTTT AGAAAGCTCT 2040

TTTTTTTTTT TTTTTTCTTT CTTTTTCTTA AGAAATAAAC GAACGTAGAT TGCGGCTCAA 2100

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 2140

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 525 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Theobroma cacao
(F) TISSUE TYPE: Seeds

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Met Val Ile Ser Lys Ser Pro Phe Ile Val Leu Ile Phe Ser Leu Leu
1 5 10 15

Leu Ser Phe Ala Leu Leu Cys Ser Gly Val Ser Ala Tyr Gly Arg Lys
20 25 30

Gln Tyr Glu Arg Asp Pro Arg Gln Gln Tyr Glu Gln Cys Gln Arg Arg
35 40 45

Cys Glu Ser Glu Ala Thr Glu Glu Arg Glu Gln Glu Gln Cys Glu Gln
50 55 60

Arg Cys Glu Arg Glu Tyr Lys Glu Gln Gln Arg Gln Gln Glu Glu Glu
65 70 75 80

Leu Gln Arg Gln Tyr Gln Gln Cys Gln Gly Arg Cys Gln Glu Gln Gln
85 90 95

Gln Gly Gln Arg Glu Gln Gln Gln Cys Gln Arg Lys Cys Trp Glu Gln
100 105 110

Tyr Lys Glu Gln Glu Arg Gly Glu His Glu Asn Tyr His Asn His Lys
115 120 125

Lys Asn Arg Ser Glu Glu Glu Glu Gly Gln Gln Arg Asn Asn Pro Tyr
130 135 140

Tyr Phe Pro Lys Arg Arg Ser Phe Gln Thr Arg Phe Arg Asp Glu Glu
145 150 155 160

Gly Asn Phe Lys Ile Leu Gln Arg Phe Ala Glu Asn Ser Pro Pro Leu
165 170 175

Lys Gly Ile Asn Asp Tyr Arg Leu Ala Met Phe Glu Ala Asn Pro Asn
180 185 190

Thr Phe Ile Leu Pro His His Cys Asp Ala Glu Ala Ile Tyr Phe Val
195 200 205

Thr Asn Gly Lys Gly Thr Ile Thr Phe Val Thr His Glu Asn Lys Glu
210 215 220

Ser Tyr Asn Val Gln Arg Gly Thr Val Val Ser Val Pro Ala Gly Ser
225 230 235 240

Thr Val Tyr Val Val Ser Gln Asp Asn Gln Glu Lys Leu Thr Ile Ala
245 250 255

Val Leu Ala Leu Pro Val Asn Ser Pro Gly Lys Tyr Glu Leu Phe Phe
260 265 270

Pro Ala Gly Asn Asn Lys Pro Glu Ser Tyr Tyr Gly Ala Phe Ser Tyr
275 280 285

Glu Val Leu Glu Thr Val Phe Asn Thr Gln Arg Glu Lys Leu Glu Glu
290 295 300

Ile Leu Glu Glu Gln Arg Gly Gln Lys Arg Gln Gln Gly Gln Gln Gly
305 310 315 320

Met Phe Arg Lys Ala Lys Pro Glu Gln Ile Arg Ala Ile Ser Gln Gln
325 330 335

Ala Thr Ser Pro Arg His Arg Gly Gly Glu Arg Leu Ala Ile Asn Leu
340 345 350

Leu Ser Gln Ser Pro Val Tyr Ser Asn Gln Asn Gly Arg Phe Phe Glu
355 360 365

Ala Cys Pro Glu Asp Phe Ser Gln Phe Gln Asn Met Asp Val Ala Val
370 375 380

Ser Ala Phe Lys Leu Asn Gln Gly Ala Ile Phe Val Pro His Tyr Asn
385 390 395 400

Ser Lys Ala Thr Phe Val Val Phe Val Thr Asp Gly Tyr Gly Tyr Ala
405 410 415

Gln Met Ala Cys Pro His Leu Ser Arg Gln Ser Gln Gly Ser Gln Ser
420 425 430

Gly Arg Gln Asp Arg Arg Glu Gln Glu Glu Glu Ser Glu Glu Glu Thr
435 440 445

Phe Gly Glu Phe Gln Gln Val Lys Ala Pro Leu Ser Pro Gly Asp Val
450 455 460

Phe Val Ala Pro Ala Gly His Ala Val Thr Phe Phe Ala Ser Lys Asp
465 470 475 480

Gln Pro Leu Asn Ala Val Ala Phe Gly Leu Asn Ala Gln Asn Asn Gln
485 490 495

Arg Ile Phe Leu Ala Gly Arg Pro Phe Phe Leu Asn His Lys Gln Asn
500 505 510

Thr Asn Val Ile Lys Phe Thr Val Lys Ala Ser Ala Tyr
515 520 525

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 590 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Gossypium hirsutum
(F) TISSUE TYPE: Seeds

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Met Val Arg Asn Lys Ser Ala Cys Val Val Leu Leu Phe Ser Leu Phe
1 5 10 15

Leu Ser Phe Gly Leu Leu Cys Ser Ala Lys Asp Phe Pro Gly Arg Arg
20 25 30

Gly Asp Asp Asp Pro Pro Lys Arg Tyr Glu Asp Cys Arg Arg Arg Cys
35 40 45

Glu Trp Asp Thr Arg Gly Gln Lys Glu Gln Gln Gln Cys Glu Glu Ser
50 55 60

Cys Lys Ser Gln Tyr Gly Glu Lys Asp Gln Gln Gln Arg His Arg Pro
65 70 75 80

Glu Asp Pro Gln Arg Arg Tyr Glu Glu Cys Gln Gln Glu Cys Arg Gln
85 90 95

Gln Glu Glu Arg Gln Gln Pro Gln Cys Gln Gln Arg Cys Leu Lys Arg
100 105 110

Phe Glu Gln Glu Gln Gln Gln Ser Gln Arg Gln Phe Gln Glu Cys Gln
115 120 125

Gln His Cys His Gln Gln Glu Gln Arg Pro Glu Lys Lys Gln Gln Cys
130 135 140

Val Arg Glu Cys Arg Glu Lys Tyr Gln Glu Asn Pro Trp Arg Gly Glu
145 150 155 160

Arg Glu Glu Glu Ala Glu Glu Glu Glu Thr Glu Glu Gly Glu Gln Glu
165 170 175

Gln Ser His Asn Pro Phe His Phe His Arg Arg Ser Phe Gln Ser Arg
180 185 190

Phe Arg Glu Glu His Gly Asn Phe Arg Val Leu Gln Arg Phe Ala Ser
195 200 205

Arg His Pro Ile Leu Arg Gly Ile Asn Glu Phe Arg Leu Ser Ile Leu
210 215 220

Glu Ala Asn Pro Asn Thr Phe Val Leu Pro His His Cys Asp Ala Glu
225 230 235 240

Lys Ile Tyr Leu Val Thr Asn Gly Arg Gly Thr Leu Thr Phe Leu Thr
245 250 255

His Glu Asn Lys Glu Ser Tyr Asn Ile Val Pro Gly Val Val Val Lys
260 265 270

Val Pro Ala Gly Ser Thr Val Tyr Leu Ala Asn Gln Asp Asn Lys Glu
275 280 285

Lys Leu Ile Ile Ala Val Leu His Arg Pro Val Asn Asn Pro Gly Gln
290 295 300

Phe Glu Glu Phe Phe Pro Ala Gly Ser Gln Arg Pro Gln Ser Tyr Leu
305 310 315 320

Arg Ala Phe Ser Arg Glu Ile Leu Glu Pro Ala Phe Asn Thr Arg Ser
325 330 335

Glu Gln Leu Asp Glu Leu Phe Gly Gly Arg Gln Ser Arg Arg Arg Gln
340 345 350

Gln Gly Gln Gly Met Phe Arg Lys Ala Ser Gln Glu Gln Ile Arg Ala

355 360 365

Leu Ser Gln Glu Ala Thr Ser Pro Arg Glu Lys Ser Gly Glu Arg Phe
370 375 380

Ala Phe Asn Leu Leu Ser Gln Thr Pro Arg Tyr Ser Asn Gln Asn Gly
385 390 395 400

Arg Phe Phe Glu Ala Cys Pro Pro Glu Phe Arg Gln Leu Arg Asp Ile
405 410 415

Asn Val Thr Val Ser Ala Leu Gln Leu Asn Gln Gly Ser Ile Phe Val
420 425 430

Pro His Tyr Asn Ser Lys Ala Thr Phe Val Ile Leu Val Thr Glu Gly
435 440 445

Asn Gly Tyr Ala Glu Met Val Ser Pro His Leu Pro Arg Gln Ser Ser
450 455 460

Tyr Glu Glu Glu Glu Glu Glu Asp Glu Glu Glu Glu Gln Glu Gln Glu
465 470 475 480

Glu Glu Arg Arg Ser Gly Gln Tyr Arg Lys Ile Arg Ser Arg Leu Ser
485 490 495

Arg Gly Asp Ile Phe Val Val Pro Ala Asn Phe Pro Val Thr Phe Val
500 505 510

Ala Ser Gln Asn Gln Asn Leu Arg Met Thr Gly Phe Gly Leu Tyr Asn
515 520 525

Gln Asn Ile Asn Pro Asp His Asn Gln Arg Ile Phe Val Ala Gly Lys
530 535 540

Ile Asn His Val Arg Gln Trp Asp Ser Gln Ala Lys Glu Leu Ala Phe
545 550 555 560

Gly Val Ser Ser Arg Leu Val Asp Glu Ile Phe Asn Ser Asn Pro Gln
565 570 575

Glu Ser Tyr Phe Val Ser Arg Gln Arg Gln Arg Ala Ser Glu
580 585 590

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

Arg Gln Arg Asp Pro Gln Gln Gln Ala Glu Gln Ala Gln Lys Arg Ala
1 5 10 15

Gln Arg Arg Glu Thr Glu
20

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Pro Arg His Met Gln Ile Ala Gln Gln Arg Ala Glu Arg Arg Ala Glu
1 5 10 15

Lys Glu Lys Arg Lys Gln Gln Lys Arg
20 25

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Met Ala Trp Phe His Val Ser Val Cys Asn Ala Val Phe Val Val Ile
1 5 10 15

Ile Ile Ile Met Leu Leu Met Phe Val Pro Val Val Arg Gly
20 25 30

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: nucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

CAGCAGCAGT ATGAGCAGTG 20

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TTTTTCGTAK CKKCKTTCGC A 21

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

ACACCATATG CGACAACGTG ATCC 24

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CGTTGTTTTC TCTATTCCTA GGGTTG 26

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met Gly His His His His His His His His His Ser Ser Gly His
1 5 10 15

Ile Glu Gly Arg His Met
20

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 90 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17

GGGAATTCCA TATGTATGAG CGTGATCCTC GACAGCAATA CGAGCAATGC CAGAGGCGAT 60

GCGAGTCGGA AGCGACTGAA GAAAGGGAGC 90

(2) INFORMATION FOR SEQ ID NO: 18

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 91 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GAAGCGACTG AAGAAAGGGA GCAAGAGCAG TGTGAACAAC GCTGTGAAAG GGAGTACAAG 60

GAGCAGCAGA GACAGCAATA GGGATCCACA C 91

(2) INFORMATION FOR SEQ ID NO: 19

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 101 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

GGGAATTCCA TATGCTTCAA AGGCAATACC AGCAATGTCA AGGGCGTTGT CAAGAGCAAC 60

AACAGGGGCA GAGAGAGCAG CAGCAGTGCC AGAGAAAATG C 101

(2) INFORMATION FOR SEQ ID NO: 20

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 102 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20

GTGTGGATCC CTAGCTCCTA TTTTTTTTGT GATTATGGTA ATTCTCGTGC TCGCCTCTCT 60

CTTGTTCCTT ATATTGCTCC CAGCATTTTC TCTGGCACTG CT 102

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Peanut
(F) TISSUE TYPE: Seeds

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

Met Arg Gly Arg Val Ser Pro Leu Met Leu Leu Leu Gly Ile Leu Val
1 5 10 15

Leu Ala Ser Val Ser Ala Thr Gln Ala Lys Ser Pro Tyr Arg Lys Thr
20 25 30

Glu Asn Pro Cys Ala Gln Arg Cys Leu Gln Ser Cys Gln Gln Glu Pro
35 40 45

Asp Asp Leu Lys Gln Lys Ala Cys Glu Ser Arg Cys Thr Lys Leu Glu
50 55 60

Tyr Asp Pro Arg Cys Val Tyr Asp Thr Gly Ala Thr Asn Gln Arg His
65 70 75 80

Pro Pro Gly Glu Arg Thr Arg Gly Arg Gln Pro Gly Asp Tyr Asp Asp
85 90 95

Asp Arg Arg Gln Pro Arg Arg Glu Glu Gly Gly Arg Trp Gly Pro Ala
100 105 110

Glu Pro Arg Glu Arg Glu Arg Glu Glu Asp Trp Arg Gln Pro Arg Glu
115 120 125

Asp Trp Arg Arg Pro Ser His Gln Gln Pro Arg Lys Ile Arg Pro Glu
130 135 140

Gly Arg Glu Gly Glu Gln Glu Trp Gly Thr Pro Gly Ser Glu Val Arg
145 150 155 160

Glu Glu Thr Ser Arg Asn Asn Pro Phe Tyr Phe Pro Ser Arg Arg Phe
165 170 175 180

Ser Thr Arg Tyr Gly Asn Gln Asn Gly Arg Ile Arg Val Leu Gln Arg
185 190 195

Phe Asp Gln Arg Ser Lys Gln Phe Gln Asn Leu Gln Asn His Arg Ile
200 205 210

Val Gln Ile Glu Ala Arg Pro Asn Thr Leu Val Leu Pro Lys His Ala
215 220 225

Asp Ala Asp Asn Ile Leu Val Ile Gln Gln Gly Gln Ala Thr Val Thr
230 235 240 245

Val Ala Asn Gly Asn Asn Arg Lys Ser Phe Asn Leu Asp Glu Gly His
250 255 260

Ala Leu Arg Ile Pro Ser Gly Phe Ile Ser Tyr Ile Leu Asn Arg His
265 270 275

Asp Asn Gln Asn Leu Arg Val Ala Lys Ile Ser Met Pro Val Asn Thr
280 285 290

Pro Gly Gln Phe Glu Asp Phe Phe Pro Ala Ser Ser Arg Asp Gln Ser
295 300 305

Ser Tyr Leu Gln Gly Phe Ser Arg Asn Thr Leu Glu Ala Ala Phe Asn
310 315 320 325

Ala Glu Phe Asn Glu Ile Arg Arg Val Leu Leu Glu Glu Asn Ala Gly
330 335 340

Gly Glu Gln Glu Glu Arg Gly Gln Arg Arg Arg Ser Thr Arg Ser Ser
345 350 355

Asp Asn Glu Gly Val Ile Val Lys Val Ser Lys Glu His Val Gln Glu
360 365 370

Leu Thr Lys His Ala Lys Ser Val Ser Lys Lys Gly Ser Glu Glu Glu
375 380 385

```
Asp Ile Thr Asn Pro Ile Asn Leu Arg Asp Gly Glu Pro Asp Leu Ser
390                 395                 400                 405

Asn Asn Phe Gly Arg Leu Phe Glu Val Lys Pro Asp Lys Lys Asn Pro
                410                 415                 420

Gln Leu Gln Asp Leu Asp Met Met Leu Thr Cys Val Glu Ile Lys Glu
            425                 430                 435

Gly Ala Leu Met Leu Pro His Phe Asn Ser Lys Ala Met Val Ile Val
        440                 445                 450

Val Val Asn Lys Gly Thr Gly Asn Leu Glu Leu Val Ala Val Arg Lys
    455                 460                 470

Glu Gln Gln Gln Arg Gly Arg Arg Glu Gln Glu Trp Glu Glu Glu Glu
480                 485                 490                 500

Glu Asp Glu Glu Glu Glu Gly Ser Asn Arg Glu Val Arg Arg Tyr Thr
                505                 510                 515

Ala Arg Leu Lys Glu Gly Asp Val Phe Ile Met Pro Ala Ala His Pro
            520                 525                 530

Val Ala Ile Asn Ala Ser Ser Glu Leu His Leu Leu Gly Phe Gly Ile
        535                 540                 545

Asn Ala Glu Asn Asn His Arg Ile Phe Leu Ala Gly Asp Lys Asp Asn
    550                 555                 560

Val Ile Asp Gln Ile Glu Lys Gln Ala Lys Asp Leu Ala Phe Pro Gly
565                 570                 575                 580

Ser Gly Glu Gln Val Glu Lys Leu Ile Lys Asn Gln Arg Glu Ser His
                585                 590                 595

Phe Val Ser Ala Arg Pro Gln Ser Gln Ser Pro Ser Ser Pro Glu Lys
            600                 605                 610

Glu Asp Gln Glu Glu Glu Asn Gln Gly Gly Lys Gly Pro Leu Leu Ser
        615                 620                 625

Ile Leu Lys Ala Phe Asn
    630
```

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 46 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Maize
(F) TISSUE TYPE: Seeds

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

Met Val Ser Ala Arg Ile Val Val Leu Leu Ala Thr Leu Leu Cys Ala
1 5 10 15

Ala Ala Ala Val Ala Ser Ser Trp Glu Asp Asp Asn His His His His
20 25 30

Gly Gly His Lys Ser Gly Gln Cys Val Arg Arg Cys Glu Asp Arg Pro
35 40 45

Trp His Gln Arg Pro Arg Cys Leu Glu Gln Cys Arg Glu Glu Glu Arg
50 55 60

Glu Lys Arg Gln Glu Arg Ser Arg His Glu Ala Asp Asp Arg Ser Gly
65 70 75 80

Glu Gly Ser Ser Glu Asp Glu Arg Glu Gln Glu Lys Glu Lys Gln Lys
85 90 95

Asp Arg Arg Pro Tyr Val Phe Asp Arg Arg Ser Phe Arg Arg Val Val
100 105 110

Arg Ser Glu Gln Gly Ser Leu Arg Val Leu Arg Pro Phe Asp Glu Val
115 120 125

Ser Arg Leu Leu Arg Gly Ile Arg Asp Tyr Arg Val Ala Val Leu Glu
130 135 140

Ala Asn Pro Arg Ser Phe Val Val Pro Ser His Thr Asp Ala His Cys
145 150 155 160

Ile Cys Tyr Val Ala Glu Gly Glu Gly Val Val Thr Thr Ile Glu Asn
165 170 175 180

Gly Glu Arg Arg Ser Tyr Thr Ile Lys Gln Gly His Val Phe Val Ala
185 190 195

Pro Ala Gly Ala Val Thr Tyr Leu Ala Asn Thr Asp Gly Arg Lys Lys
200 205 210

Leu Val Ile Thr Lys Ile Leu His Thr Ile Ser Val Pro Gly Glu Phe
215 220 225

Gln Phe Phe Phe Gly Pro Gly Gly Arg Asn Pro Glu Ser Phe Leu Ser
230 235 240 245

Ser Phe Ser Lys Ser Ile Gln Arg Ala Ala Tyr Lys Thr Ser Ser Asp
250 255 260

Arg Leu Glu Arg Leu Phe Gly Arg His Gly Gln Asp Lys Gly Ile Ile

265 270 275

Val Arg Ala Thr Glu Glu Gln Thr Arg Glu Leu Arg Arg His Ala Ser
280 285 290

Glu Gly Gly His Gly Pro His Trp Pro Leu Pro Pro Phe Gly Glu Ser
295 300 305

Arg Gly Pro Tyr Ser Leu Leu Asp Gln Arg Pro Ser Ile Ala Asn Gln
310 315 320 325

His Gly Gln Leu Tyr Glu Ala Asp Ala Arg Ser Phe His Asp Leu Ala
330 335 340

Glu His Asp Val Ser Val Ser Phe Ala Asn Ile Thr Ala Gly Ser Met
345 350 355

Ser Ala Pro Leu Phe Asn Thr Arg Ser Phe Lys Ile Ala Tyr Val Pro
360 365 370

Asn Gly Lys Gly Tyr Ala Glu Ile Val Cys Pro His Arg Gln Ser Gln
375 380 385

Gly Gly Glu Ser Glu Arg Glu Arg Asp Lys Gly Arg Arg Ser Glu Glu
390 395 400 405

Glu Glu Glu Glu Ser Ser Glu Glu Gln Glu Glu Ala Gly Gln Gly Tyr
410 415 420

His Thr Ile Arg Ala Arg Leu Ser Pro Gly Thr Ala Phe Val Val Pro
425 430 435

Ala Gly His Pro Phe Val Ala Val Ala Ser Arg Asp Ser Asn Leu Gln
440 445 450

Ile Val Cys Phe Glu Val His Ala Asp Arg Asn Glu Lys Val Phe Leu
455 460 470

Ala Gly Ala Asp Asn Val Leu Gln Lys Leu Asp Arg Val Ala Lys Ala
480 485 490 500

Leu Ser Phe Ala Ser Lys Ala Glu Glu Val Asp Glu Val Leu Gly Ser
505 510 515

Arg Arg Glu Lys Gly Phe Leu Pro Gly Pro Glu Glu Ser Gly Gly His
520 525 530

Glu Glu Arg Glu Gln Glu Glu Glu Glu Arg Glu Glu Arg His Gly Gly
535 540 545

Arg Gly Glu Arg Glu Arg His Gly Arg Glu Glu Arg Glu Lys Glu Glu
550 555 560

Glu Arg Glu Gly Arg His Gly Gly Arg Glu Glu Arg Glu Glu Glu Glu
565 570 575 580

Arg His Gly Arg Gly Arg Arg Glu Glu Val Ala Glu Thr Leu Met Arg
585 590 595

Met Val Thr Ala Arg Met
600

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Maize
(F) TISSUE TYPE: Seeds

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

Arg Ser Gly Arg Gly Glu Cys Arg Arg Gln Cys Leu Arg Arg His Glu
1 5 10 15

Gly Gln Pro Trp Glu Thr Gln Glu Cys Met Arg Arg Cys Arg Arg Arg
20 25 30

Gly

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Barley
(F) TISSUE TYPE: Seeds

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

Met Ala Thr Arg Ala Lys Ala Thr Ile Pro Leu Leu Phe Leu Leu Gly
1 5 10 15

Thr Ser Leu Leu Phe Ala Ala Ala Val Ser Ala Ser His Asp Asp Glu
20 25 30

Asp Asp Arg Arg Gly Gly His Ser Leu Gln Gln Cys Val Gln Arg Cys
35 40 45

Arg Gln Glu Arg Pro Arg Tyr Ser His Ala Arg Cys Val Gln Glu Cys
50 55 60

Arg Asp Asp Gln Gln Gln His Gly Arg His Glu Gln Glu Glu Glu Gln
65 70 75 80

Gly Arg Gly Arg Gly Trp His Gly Glu Gly Glu Arg Glu Glu Glu His
85 90 95

Gly Arg Gly Arg Gly Arg His Gly Glu Gly Glu Arg Glu Glu Glu His
100 105 110

Gly Arg Gly Arg Gly Arg His Gly Glu Gly Glu Arg Glu Glu Glu Arg
115 120 125

Gly Arg Gly His Gly Arg His Gly Glu Gly Glu Arg Glu Glu Glu Arg
130 135 140

Gly Arg Gly Arg Gly Arg His Gly Glu Gly Glu Arg Glu Glu Glu Glu
145 150 155 160

Gly Arg Gly Arg Gly Arg Arg Gly Glu Gly Glu Arg Asp Glu Glu Gln
165 170 175 180

Gly Asp Ser Arg Arg Pro Tyr Val Phe Gly Pro Arg Ser Phe Arg Arg
185 190 195

Ile Ile Gln Ser Asp His Gly Phe Val Arg Ala Leu Arg Pro Phe Asp
200 205 210

Gln Val Ser Arg Leu Leu Arg Gly Ile Arg Asp Tyr Arg Val Ala Ile
215 220 225

Met Glu Val Asn Pro Arg Ala Phe Val Val Pro Gly Phe Thr Asp Ala
230 235 240 245

Asp Gly Val Gly Tyr Val Ala Gln Gly Glu Gly Val Leu Thr Val Ile
250 255 260

Glu Asn Gly Glu Lys Arg Ser Tyr Thr Val Lys Glu Gly Asp Val Ile
265 270 275

Val Ala Pro Ala Gly Ser Ile Met His Leu Ala Asn Thr Asp Gly Arg
280 285 290

Arg Lys Leu Val Ile Ala Lys Ile Leu His Thr Ile Ser Val Pro Gly
295 300 305

Lys Phe Gln Phe Leu Ser Val Lys Pro Leu Leu Ala Ser Leu Ser Lys
310 315 320 325

Arg Val Leu Arg Ala Ala Phe Lys Thr Ser Asp Glu Arg Leu Glu Arg
330 335 340

Leu Phe Asn Gln Arg Gln Gly Gln Glu Lys Thr Arg Ser Val Ser Ile

345 350 355

Val Arg Ala Ser Glu Glu Gln Leu Arg Glu Leu Arg Arg Glu Ala Ala
360 365 370

Glu Gly Gly Gln Gly His Arg Trp Pro Leu Pro Pro Phe Arg Gly Asp
375 380 385

Ser Arg Asp Thr Phe Asn Leu Leu Glu Gln Arg Pro Lys Ile Ala Asn
390 395 400 405

Arg His Gly Arg Leu Tyr Glu Ala Asp Ala Arg Ser Phe His Ala Leu
410 415 420

Ala Asn Gln Asp Val Arg Val Ala Val Ala Asn Ile Thr Pro Gly Ser
425 430 435

Met Thr Ala Pro Tyr Leu Asn Thr Gln Ser Phe Lys Leu Ala Val Val
440 445 450

Leu Glu Gly Glu Gly Glu Val Gln Ile Val Cys Pro His Leu Gly Arg
455 460 470

Glu Ser Glu Ser Glu Arg Glu His Gly Lys Gly Arg Arg Arg Glu Glu
480 485 490 500

Glu Glu Asp Asp Gln Arg Gln Gln Arg Arg Arg Gly Ser Glu Ser Glu
505 510 515

Ser Glu Glu Glu Glu Glu Gln Gln Arg Tyr Glu Thr Val Arg Ala Arg
520 525 530

Val Ser Arg Gly Ser Ala Phe Val Val Pro Pro Gly His Pro Val Val
535 540 545

Glu Ile Ser Ser Ser Gln Gly Ser Ser Asn Leu Gln Val Val Cys Phe
550 555 560

Glu Ile Asn Ala Glu Arg Asn Glu Arg Val Trp Leu Ala Gly Arg Asn
565 570 575 580

Asn Val Ile Gly Lys Leu Gly Ser Pro Ala Gln Glu Leu Thr Phe Gly
585 590 595

Arg Pro Ala Arg Glu Val Gln Glu Val Phe Arg Ala Glu Asp Glu Asp
600 605 610

Glu Gly Phe Val Ala Gly Pro Glu Gln Gln Ser Arg Glu Gln Glu Gln
615 620 625

Glu Gln Glu Arg His Arg Arg Arg Gly Asp Arg Gly Arg Gly Asp Glu
630 635 640

Ala Val Glu Thr Phe Leu Arg Met Ala Thr Gly Ala Ile
645 650 655

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 55 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Soybean (Glycine max)
(F) TISSUE TYPE: Seeds

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

Met Met Arg Ala Arg Phe Pro Leu Leu Leu Leu Gly Leu Val Phe Leu
1 5 10 15

Ala Ser Val Ser Val Ser Phe Gly Ile Ala Tyr Trp Glu Lys Glu Asn
20 25 30

Pro Lys His Asn Lys Cys Leu Gln Ser Cys Asn Ser Glu Arg Asp Ser
35 40 45

Tyr Arg Asn Gln Ala Cys His Ala Arg Cys Asn Leu Leu Lys Val Glu
50 55 60

Lys Glu Glu Cys Glu Glu Gly Glu Ile Pro Arg Pro Arg Pro Arg Pro
65 70 75 80

Gln His Pro Glu Arg Glu Pro Gln Gln Pro Gly Glu Lys Glu Glu Asp
85 90 95

Glu Asp Glu Gln Pro Arg Pro Ile Pro Phe Pro Arg Pro Gln Pro Arg
100 105 110

Gln Glu Glu Glu His Glu Gln Arg Glu Glu Gln Glu Trp Pro Arg Lys
115 120 125

Glu Glu Lys Arg Gly Glu Lys Gly Ser Glu Glu Glu Asp Glu Asp Glu
130 135 140

Asp Glu Glu Gln Asp Glu Arg Gln Phe Pro Phe Pro Arg Pro Pro His
145 150 155 160

Gln Lys Glu Glu Arg Asn Glu Glu Glu Asp Glu Asp Glu Glu Gln Gln
165 170 175 180

Arg Glu Ser Glu Glu Ser Glu Asp Ser Glu Leu Arg Arg His Lys Asn
185 190 195

Lys Asn Pro Phe Leu Phe Gly Ser Asn Arg Phe Glu Thr Leu Phe Lys
200 205 210

Asn Gln Tyr Gly Arg Ile Arg Val Leu Gln Arg Phe Asn Gln Arg Ser
215 220 225

Pro Gln Leu Gln Asn Leu Arg Asp Tyr Arg Ile Leu Glu Phe Asn Ser
230 235 240 245

Lys Pro Asn Thr Leu Leu Leu Pro Asn His Ala Asp Ala Asp Tyr Leu
250 255 260

Ile Val Ile Leu Asn Gly Thr Ala Ile Leu Ser Leu Val Asn Asn Asp
265 270 275

Asp Arg Asp Ser Tyr Arg Leu Gln Ser Gly Asp Ala Leu Arg Val Pro
280 285 290

Ser Gly Thr Thr Tyr Tyr Val Val Asn Pro Asp Asn Asn Glu Asn Leu
295 300 305

Arg Leu Ile Thr Leu Ala Ile Pro Val Asn Lys Pro Gly Arg Phe Glu
310 315 320 325

Ser Phe Phe Leu Ser Ser Thr Glu Ala Gln Gln Ser Tyr Leu Gln Gly
330 335 340

Phe Ser Arg Asn Ile Leu Glu Ala Ser Tyr Asp Thr Lys Phe Glu Glu
345 350 355

Ile Asn Lys Val Leu Phe Ser Arg Glu Glu Gly Gln Gln Gln Gly Glu
360 365 370

Gln Arg Leu Gln Glu Ser Val Ile Val Glu Ile Ser Lys Glu Gln Ile
375 380 385

Arg Ala Leu Ser Lys Arg Ala Lys Ser Ser Ser Arg Lys Thr Ile Ser
390 395 400 405

Ser Glu Asp Lys Pro Phe Asn Leu Arg Ser Arg Asp Pro Ile Tyr Ser
410 415 420

Asn Lys Leu Gly Lys Phe Phe Glu Ile Thr Pro Glu Lys Asn Pro Gln
425 430 435

Leu Arg Asp Leu Asp Ile Phe Leu Ser Ile Val Asp Met Asn Glu Gly
440 445 450

Ala Leu Leu Leu Pro His Phe Asn Ser Lys Ala Ile Val Ile Leu Val
455 460 470

Ile Asn Glu Gly Asp Ala Asn Ile Glu Leu Val Gly Leu Lys Glu Gln
480 485 490 500

Gln Gln Glu Gln Gln Gln Glu Glu Gln Pro Leu Glu Val Arg Lys Tyr
505 510 515

Arg Ala Glu Leu Ser Glu Gln Asp Ile Phe Val Ile Pro Ala Gly Tyr
520 525 530

```
Pro Val Val Val Asn Ala Thr Ser Asn Leu Asn Phe Phe Ala Ile Gly
        535                 540                 545

Ile Asn Ala Glu Asn Asn Gln Arg Asn Phe Leu Ala Gly Ser Gln Asp
    550                 555                 560

Asn Val Ile Ser Gln Ile Pro Ser Gln Val Gln Glu Leu Ala Phe Pro
565                 570                 575                 580

Gly Ser Ala Gln Ala Val Glu Lys Leu Leu Lys Asn Gln Arg Glu Ser
                585                 590                 595

Tyr Phe Val Asp Ala Gln Pro Lys Lys Lys Glu Glu Gly Asn Lys Gly
            600                 605                 610

Arg Lys Gly Pro Leu Ser Ser Ile Leu Arg Ala Phe Tyr
        615                 620                 625
```

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Stenocarpus sinuatus
(F) TISSUE TYPE: Seeds

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

```
Val Lys Glu Asp His Gln Phe Glu Thr Arg Gly Glu Ile Leu Glu Cys
1               5                   10                  15

Tyr Arg Leu Cys Gln Gln Gln
            20
```

(28) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Stenocarpus sinuatus
(F) TISSUE TYPE: Seeds

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

Gln Lys His Arg Ser Gln Ile Leu Gly Cys Tyr Leu Xxx cys Gln Gln
1 5 10 15

Leu

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Stenocarpus sinuatus
(F) TISSUE TYPE: Seeds

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

Leu Asp Pro Ile Arg Gln Gln Gln Leu Cys Gln Met Arg Cys Gln Gln
1 5 10 15

Gln Glu Lys Asp Pro Arg Gln Gln Gln Gln Cys Lys
20 25

CLAIMS

1. A protein fragment having antimicrobial activity, wherein said protein fragment is selected from:

(ii) a polypeptide having an amino acid sequence selected from:

residues 29 to 73 of SEQ ID NO: 1
residues 74 to 116 of SEQ ID NO: 1
residues 117 to 185 of SEQ ID NO: 1
residues 186 to 248 of SEQ ID NO: 1
residues 29 to 73 of SEQ ID NO: 3
residues 74 to 116 of SEQ ID NO: 3
residues 117 to 185 of SEQ ID NO: 3
residues 186 to 248 of SEQ ID NO: 3
residues 1 to 32 of SEQ ID NO: 5
residues 33 to 75 of SEQ ID NO: 5
residues 76 to 144 of SEQ ID NO: 5
residues 145 to 210 of SEQ ID NO: 5
residues 34 to 80 of SEQ ID NO: 7
residues 81 to 140 of SEQ ID NO: 7
residues 33 to 79 of SEQ ID NO: 8
residues 80 to 119 of SEQ ID NO: 8
residues 120 to 161 of SEQ ID NO: 8
residues 32 to 91 of SEQ ID NO: 21
residues 25 to 84 of SEQ ID NO: 22
residues 29 to 94 of SEQ ID NO: 24
residues 31 to 85 of SEQ ID NO: 25
residues 1 to 23 of SEQ ID NO: 26
residues 1 to 17 of SEQ ID NO: 27
residues 1 to 28 of SEQ ID NO: 28;

(ii) a homologue of (i);

(iii) a polypeptide containing a relative cysteine spacing of C-2X-C-3X-C-(10-12)X-C-3X-C-3X-C wherein X is any amino acid residue, and C is cysteine;

(iv) a polypeptide containing a relative cysteine and tyrosine/phenylalanine spacing of Z-2X-C-3X-C-(10-12)X-C-3X-C-3X-Z wherein X is any amino acid residue, and C is cysteine, and Z is tyrosine or phenylalanine;

(v) a polypeptide containing a relative cysteine spacing of C-3X-C-(10-12)X-C-3X-C wherein X is any amino acid residue, and C is cysteine;

(vi) a polypeptide with substantially the same spacing of positively charged residues relative to the spacing of cysteine residues as (i); and

(vii) a fragment of the polypeptide of any one of (i) to (vi) which has substantially the same antimicrobial activity as (i).

2. A protein containing at least one polypeptide fragment according to claim 1, wherein said polypeptide fragment has a sequence selected from within a sequence comprising SEQ ID NO: 1, SEQ ID NO: 3 or SEQ ID NO: 5

3. A protein having a sequence selected from SEQ ID NO: 1, SEQ ID NO: 3 or SEQ ID NO: 5.

4. An isolated or synthetic DNA encoding a polypeptide fragment according to claim 1.

5. The DNA according to claim 4, wherein said DNA has a sequence selected from SEQ ID NO: 2, SEQ ID NO: 4 or SEQ ID NO: 6.

6. A DNA construct which includes a DNA according to claim 4 operatively linked to elements for the expression of said encoded protein.

7. A transgenic plant harbouring a DNA construct according to claim 6.

8. The transgenic plant according to claim 7, wherein said plant is a monocotyledonous plant or a dicotyledonous plant.

9. The transgenic plant according to claim 7, wherein said plant is selected from maize, banana, peanut, field peas, sunflower, tomato, canola, tobacco, wheat, barley, oats, potato, soybeans, cotton, carnations, roses, or sorghum.

10. Reproductive material of a transgenic plant according to claim 7.

11. A composition comprising an antimicrobial protein according to claim 1 together with an agriculturally-acceptable carrier diluent or excipient.

12. A composition comprising an antimicrobial protein according to claim1 together with an pharmaceutically-acceptable carrier diluent or excipient.

13. A method of controlling microbial infestation of a plant, the method comprising:

i) treating said plant with an antimicrobial protein according to claim 1 or a composition according to claim 11; or

ii) introducing a DNA construct according to claim 6 into said plant.

14. A method of controlling microbial infestation of a mammalian animal, the method comprising treating the animal with an antimicrobial protein according to claim 1 or a composition according to claim 12.

15. The method of claim 14, wherein said mammalian animal is a human.

16. A method of preparing an antimicrobial protein, which method comprises the steps of:

a) obtaining or designing an amino acid sequence which forms a helix-turn-helix structure;

b) replacing individual residues to achieve substantially the same distribution of positively charged residues and cysteine residues as in one or more of the amino acid sequences shown in Figure 4;

c) synthesising a protein comprising said amino acid sequence chemically or by recombinant DNA techniques in liquid culture; and

d) if necessary, forming disulphide linkages between said cysteine residues.

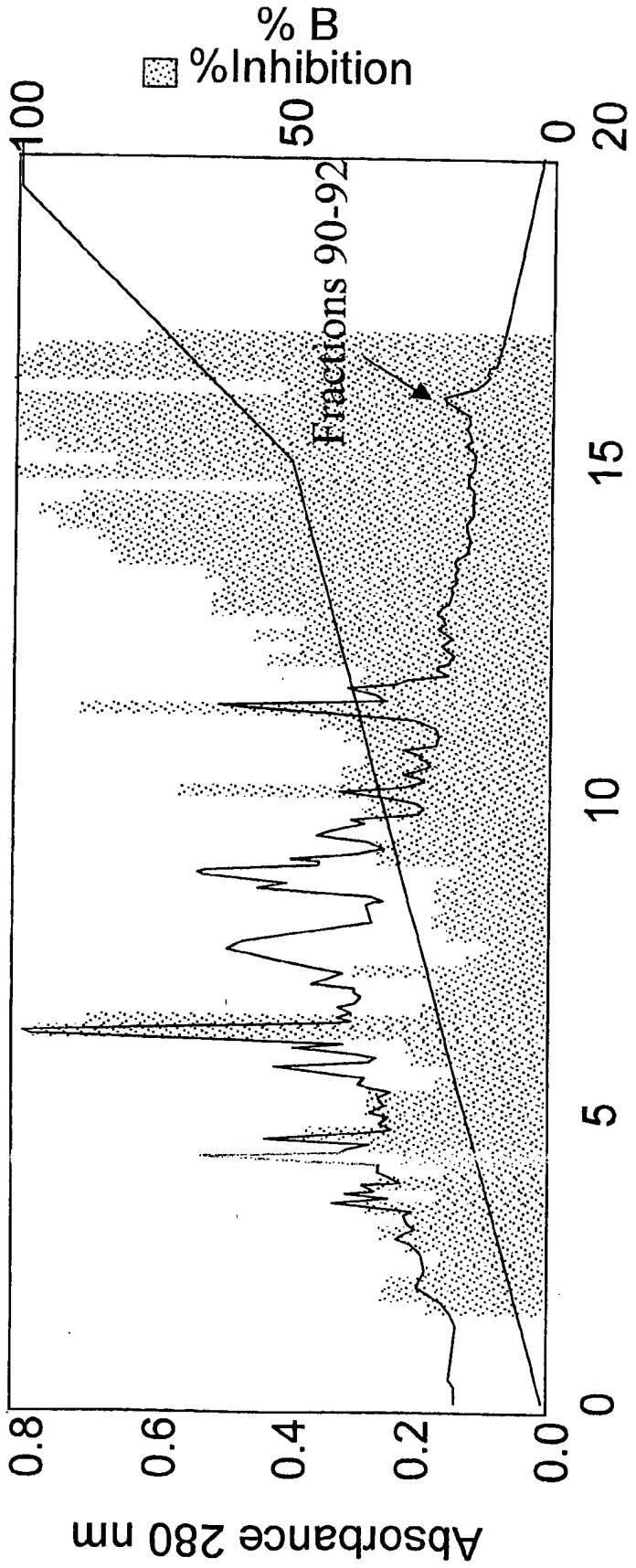
% B
%Inhibition
100
50
0
20
15
10
5
0
Hours
Fractions 90-92
0.8
0.6
0.4
0.2
0.0
Absorbance 280 nm

Fig. 1

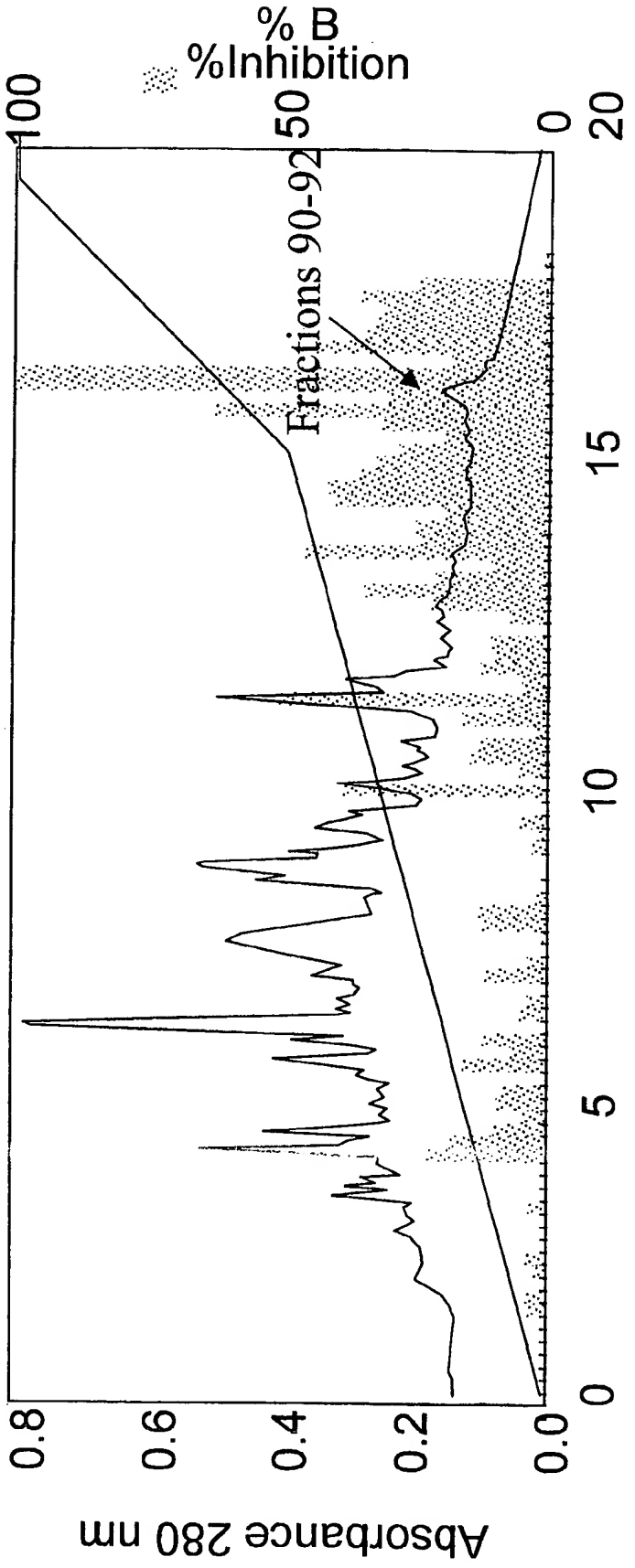
% B
%Inhibition
100
50
0
20
15
10
5
0
Hours
Fractions 90-92
0.8
0.6
0.4
0.2
0.0
Absorbance 280 nm

Fig. 2

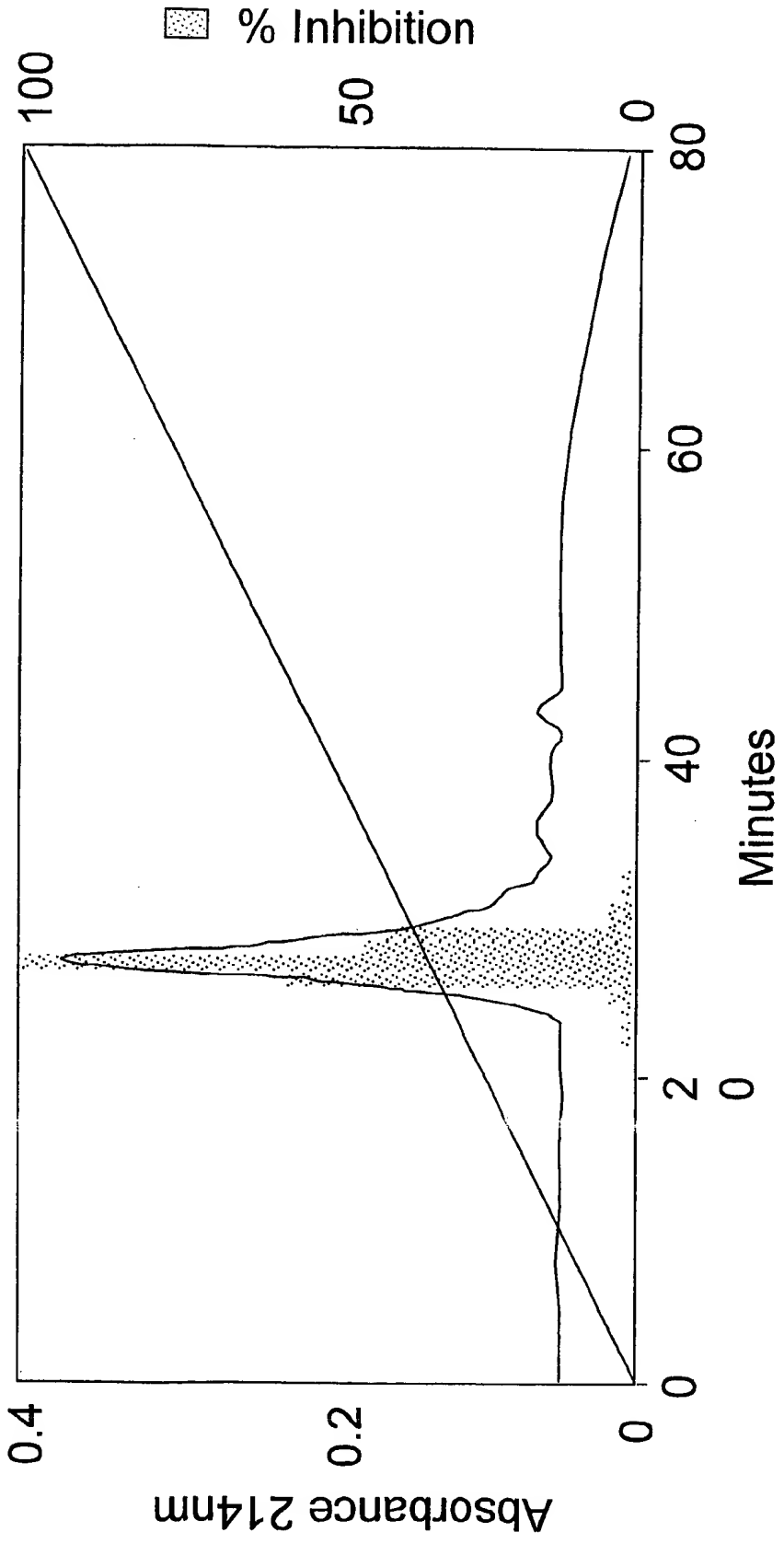
% Inhibition
100
50
0
80
60
40
20
0
Minutes
0.4
0.2
0
Absorbance 214nm

Fig. 3

```
Mi2a                1        SEFDRQEYEECKRQCMQLE-TSG-QMRRCVSQCD 32
Mi2b                1      NQEDPQTECQQCQRRCRQQE-SGPRQQQYCQRRCK 34
Mi2c                1     NRQRDPQQQYEQCQKHCQRRE-TEPRHMQTCQQRCE 35
Mi2d                1      KRDPQQREYEDCRRRCEQQE---PRQQHQCQLRCR 32
Cocoa-a             1      YERDPRQQYEQCQRRCESEA-TEEREQEQCEQRCE 34
Cocoa-b             1         LQRQYQQCQGRCQEQQ-QGQREQQQCQRKCW 30
Cotton-a            1      GDDDPPKRYEDCRRRCEWDT-RGQKEQQQCEESCK 34
Cotton-b            1       PEDPQRRYEECQQECRQQE---ERQQPQCQQRCL 31
Cotton-c            1         SQRQFQECQQHCHQQE-QRPEKKQQCVRECR 30
maize glb1_0 fr     1  EDDNHHHHGGHKSGRCVRRCEDR---PWHQRPRCLEQCR 36
barley glob fra     1  HDDEDDRRGGHSLQQCVQRCRQER--PRYSHARCVQECR 37
Peanut-a            1       TENP--CAQRCLQSCQQE--PDDLKQKACESRCT 30
alpha conglycin     1        ENP--KHNKCLQSCNSER--DSYRNQACHARCN 29
SsAMP1 partial      1  VKEDHQFETRGEILECYRLCQQQ                  23
SsAMP2 partial      1        QKHRSQILGCYLXCQQL                  17
SsAMP3 partial      1        LDPIRQQQLCQMRCQQQEKD-PRQQQQCK      28
```

Fig. 4(1/2)

Mi2a	33	KRFEEDIDWSKYD	45
Mi2b	35	EICEEEEEY	43
Mi2c	36	RRYEKEKRKQQKRYEEQQREDEEKYEERMKEEDN	69
Mi2d	33	EQQRQHGRGGDMMNPQRGGSGRYEEGEEEQS	63
Cocoa-a	35	REYKEQQRQQEEE	47
Cocoa-b	31	EQYKEQERGEHENYHNHKKNRSEEEEGQQR	60
Cotton-a	35	SQYGEKDQQQRHR	47
Cotton-b	32	KRFEQEQQQ	40
Cotton-c	31	EKYQENPWRGER	42
maize glb1	37	EEEREKRQERSRHEADDRSGEGSS	60
barley glob	38	DDQQQHGRHEQEEEQGRGRGWHGEGEREE	66
Peanut-a	31	KLEYDPRCVYDTGATNQRHPPGERT--RGRQP	60
alpha conglycin	30	LLKVEKEECEEGEIPRPRPRPQHPER	55
SsAMP1 partial	23		23
SsAMP2 partial	17		17
SsAMP3 partial	28		28

Fig. 4 (2/2)

```
AACTCTAGAG CGGCCGCGTC GACTATTTTT ACAACAATTA CCAACAACAA CAAACAACAA 60

ACAACATTAC AATTACTATT TACAATTACA GGATCCACAA CAATGGCTTG GTTCCACGTT 120
                                            M  A  W     F  H  V>

TCTGTTTGTA ACGCTGTTTT CGTTGTTATT ATTATTATTA TGCTTCTTAT GTTCGTTCCT 180
 S  V  C    N  A  V  F    V  V  I    I  I  I    M  L  L  M    F  V  P>

GTTGTTAGAG GTAGACAAAG AGATCCTCAA CAACAATACG AGCAATGTCA AAAGAGGTGT 210
 V  V  R    G  R  Q  R    D  P  Q    Q  Q  Y    E  Q  C  Q    K  R  C>
            A

CAAAGGAGAG AGACTGAGCC TAGACACATG CAAATTTGTC AGCAAAGGTG TGAAAGGAGG 240
 Q  R  R    E  T  E  P    R  H  M    Q  I  C    Q  Q  R  C    E  R  R>

TACGAGAAGG AGAAGAGGAA GCAACAAAAG AGGTGAGGAT CCGTCGACGC GGCCGCAGAT 270
 Y  E  K    E  K  R  K    Q  Q  K    R  *

CTAGACAA 278
```

Fig. 5

```
Mi clone 1          1   MAINTSNLCSLLFLLSL-FLLSTTVSLAE----SEFDRQEYEE  38
Mi clone 2          1   MAINTSNLCSLLFLLSL-FLLSTTVSLAE----SEFDRQEYEE  38
Mi clone 3          0                                                0
cotton vicilin      1   MVRNKSACVVLLFSLFLSFGLLCSAKDFPGRRGDD--------  35
cocoa vicilin       1   MVISKSPFIVLIFSLLLSFALLCSGVSAYGRKQYER-------  36
                        *. . *    *.* * * *.*  ..  .

Mi clone 1         39 CKRQCMQLETSGQMRRCVSQCDKRFEEDIDWSKYDNQEDPQTECQ  83
Mi clone 2         39 CKRQCMQLETSGQMRRCVSQCDKRFEEDIDWSKYDNQdDPQTdCQ  83
Mi clone 3         42    QCMQLETSGQMRRCVSQCDKRFEEDIDWSKYDNQEDPQTECQ  83
cotton vicilin     36 ------------------------------------DPPKRYE  42
cocoa vicilin      37 ------------------------------------DPRQQYE  43
                                                          **    .

Mi clone 1         84 QCQRRCRQQESGPRQQQYCQRRCKEICEEEEEYNRQR--DPQQQY 126
Mi clone 2         84 QCQRRCRQQESGPRQQQYCQRRCKEICEEEEEYNRQR--DPQQQY 126
Mi clone 3         84 QCQRRCRQQESdPRQQQYCQRRCKEICEEEEEYNRQR--DPQQQY 126
cotton vicilin     43 DCRRRCEWDTRGQKEQQQCEESCKSQYGEKDQQQRHRPEDPQRRY  87
cocoa vicilin      44 QCQRRCESEATEEREQEQCEQRCEREYKEQQRQQ---EEELQRQY  85
                       *.***       ..*. *.  *    *    .     . *..*
```

Fig. 6 (1/6)

```
Mi clone 1      127 EQCQKhCQRRETEPRHMQTCQQRCERRYEKEKRKQQKRYEEQQRE 171
Mi clone 2      127 EQCQeRCQRhETEPRHMQTCQQRCERRYEKEKRKQQKRYEEQQRE 171
Mi clone 3      127 EQCQKRCQRRETEPRHMQICQQRCERRYEKEKRKQQKRYEEQQRE 171
cotton vicilin   88 EECQQECRQQEE--RQQPQCQQRCLKRFEQEQQ------------ 118
cocoa vicilin    86 QQCQGRCQEQQQGQREQQQCQRKCWEQY-KEQ------------ 116
                    ..**  *. ..   *    **..*  .. .*.

Mi clone 1      172 DEEKYEERMKEEDNKRDPQQREYEDCRRRCEQQE--PRQQHQCQl 214
Mi clone 2      172 DEEKYEERMKEEDNKRDPQQREYEDCRRRCEQQE--PRQQYQCQR 214
Mi clone 3      172 DEEKYEERMKEgDNKRDPQQREYEDCRRhCEQQE--PRlQYQCQR 214
cotton vicilin 119 -----------------QSQRQFQECQQHCHQQEQRPEKKQQCVR 146
cocoa vicilin  117 ---------------------------------------------- 116

Mi clone 1      215 RCREQQRQHGRGGDmMNPQRGGSGRYEEGEEeQSDNPYYF-DERS 258
Mi clone 2      215 RCREQQRQHGRGGDLiNPQRGGSGRYEEGEEKQSDNPYYF-DERS 258
Mi clone 3      215 RCqEQQRQHGRGGDLMNPQRGGSGRYEEGEEKQSDNPYYF-DERS 258
cotton vicilin 147 ECREKY--QENPWRGEREEEAEEEETEEGEQEQSHNPFHF-HRRS 188
cocoa vicilin  117 ----------ER-GEHENYHNHKKNRSEEEEGQQRNNPYYFPKRRS 151
                                          ** *  *  **. *   **
```

Fig. 6 (2/6)

```
Mi clone 1      259 LSTRFRTEEGHISVLENFYGRSKLLRALKNYRLVLLEANPNAFVL 303
Mi clone 2      259 LSTRFRTEEGHISVLENFYGRSKLLRALKNYRLVLLEANPNAFVL 303
Mi clone 3      259 LSTRFRTEEGHISVLENFYGRSKLLRALKNYRLVLLEANPNAFVL 303
cotton vicilin  189 FQSRFREEHGNFRVLQRFASRHPILRGINEFRLSILEANPNTFVL 233
cocoa vicilin   152 FQTRFRDEEGNFKILQRFAENSPPLKGINDYRLAMFEANPNTFIL 196
                      .*** * *.  .*. *     *. .  .** . *****.*.*

Mi clone 1      304 PTHLDADAILLVIGGRGALKMIHhDNRESYNLECGDVIRIPAGTT 348
Mi clone 2      304 PTHLDADAILLVTGGRGALKMIHRDNRESYNLECGDVIRIPAGTT 348
Mi clone 3      304 PTHLDADAILLVIGGRGALKMIHRDNRESYNLECGDVIRIPAGTT 348
cotton vicilin  234 PHHCDAEKIYLVTNGRGTLTFLTHENKESYNIVPGVVVKVPAGST 278
cocoa vicilin   197 PHHCDAEAIYFVTNGKGTITFVTHENKESYNVQRGTVVSVPAGST 241
                    * * **. *  * .*.*..  . ..*.****.  * *. .***.*

Mi clone 1      349 FYLINRDNNERLHIAKFLQTISTPGQYKEFFPAGGQNPEPYLSTF 393
Mi clone 2      349 FYLINRDNNERLHIAKFLQTISTPGQYKEFFPAGGQNPEPYLSTF 393
Mi clone 3      349 FYLINRDNNERLHIAKFLQTISTPGQYKEFFPAGGQNPEPYLSTF 393
cotton vicilin  279 VYLANQDNKEKLIIAVLHRPVNNPGQFEEFFPAGSQRPQSYLRAF 323
cocoa vicilin   242 VYVVSQDNQEKLTIAVLALPVNSPGKYELFFPAGNNKPESYYGAF 286
                     *.  .** *.* **     .  **..  ***** . *. *  .*
```

Fig. 6 (3/6)

```
Mi clone 1      394 SKEILEAALNTQTEkLRGVf----GQQRE-GVIIRASQEQIRELT 433
Mi clone 2      394 SKEILEAALNTQaERLRGVL----GQQRE-GVIISASQEQIRELT 433
Mi clone 3      394 SKEILEAALNTQTERLRGVL----GQQRE-GVIIRASQEQIRELT 433
cotton vicilin 324 SREILEPAFNTRSEQLDELFGGRQSRRRQQGQG-MFRKASQEQIR 367
cocoa vicilin  287 SYEVLETVFNTQREKLEEILEEQRGQKRQQGQQGMFRKAKPEQIR 331
                   * *.**   **. *.*  .      ..*. *      .    ..

Mi clone 1      434 RDDSESRhWHIRRGGESSRGPYNLFNKRPLYSNKYGQAYEVKPED 478
Mi clone 2      434 RDDSESRRWHIRRGGESSRGPYNLFNKRPLYSNKYGQAYEVKPED 478
Mi clone 3      434 RDDSESRRWHIRRGGESSRGPYNLFNKRPLYSNKYGQAYEVKPED 478
cotton vicilin 368 ALSQEATSPREK-SGE--RFAFNLLSQTPRYSNQNGRFFEACPPE 409
cocoa vicilin  332 AISQQATSPRHR-GGE--RLAINLLSQSPVYSNQNGRFFEACPED 373
                      ..   . .  **  *   **  . * ***. *. .*  * .

Mi clone 1      479 YRQLQDMDlSVFIANvTQGSMMGPFFNTRSTKVVVVASGEADVEM 523
Mi clone 2      479 YRQLQDMDVSVFIANITQGSMMGPFFNTRSTKVVVVASGEADVEM 523
Mi clone 3      479 YRQLQDMDVSVFIANITQGSMMGPFFNTRSTKVVVVASGEADVEM 523
cotton vicilin 410 FRQLRDINVTVSALQLNQGSIFVPHYNSKATFVILVTEGNGYAEM 454
cocoa vicilin  374 FSQFQNMDVAVSAFKLNQGAIFVPHYNSKATFVVFVTDGYGYAQM 418
                   . * . . ..*    ..**..  * .*...* *. *. *    .*
```

Fig. 6 (4/6)

```
Mi clone 1      524 ACPHLSGRHGGRGGGKRHEEEED-----VHYEQVRARLSKREAIV 563
Mi clone 2      524 ACPHLSGRHGGRrGGKRHEEEED-----VHYEQVkARLSKREAIV 563
Mi clone 3      524 ACPHLSGRHGGRGGGKRHEEEEE-----VHYEQVRARLSKREAIV 563
cotton vicilin 455 VSPHLPRQSSYEEEEEEDEEEEQEQEEERRSGQYRKIRSRLSRGD 499
cocoa vicilin  419 ACPHLSRQSQGSQSGRQDRREQEEESEEETFGEFQQVKAPLSPGD 463
                    .***  .            *.          . .   .

Mi clone 1      564 ---VLAGHPVVFVSSGNENLLLFAFGINAQNNHEN-----FLAGR 600
Mi clone 2      564 ---VpvGHPVVFVSSGNENLLLFAFGINAQNNHEN-----FLAGR 600
Mi clone 3      564 ---VLAGHPVVFVSSGNENLLLFAFGINAQNNHEN-----FLAGR 600
cotton vicilin 500 IFVVPANFPVTFVASQNQNLRMTGFGLYNQNINPDHNQRIFVAGK 544
cocoa vicilin  464 VFVAPAGHAVTFFASKDQPLNAVAFGLNAQN-----NQRIFLAGR 503
                      .  * * .*  . *    **.  **         *.**.

Mi clone 1      601 ERNVLQQIEPQAMELAFAAPRKEVEEsFNSQ-DqSIFFPGPRQHQQ 645
Mi clone 2      601 ERNVLQQIEPQAMELAFAAPRKEVEELFNSQ-DESIFFPGPRQHQQ 645
Mi clone 3      601 ERNVLQQIEPQAMELAFAAsRKEVEELFNSQ-DESIFFPGPRQHQQ 645
cotton vicili 545 INHVRQ-WDSQAKELAFGVSSRLVDEIFNSNPQES-YF-VSRQRQR 587
cocoa vicilin 504 --------- ----------------------PFFLNHKQNTN 514
                                                  .*  ..
```

Fig. 6 (5/6)

```
Mi clone 1          646 QSPRSTKQQQPLVSILDFVGF 666
Mi clone 2          646 QSSRSTKQQQPLVSILDFVGF 666
Mi clone 3          646 QSPRSTKQQQPLVSILDFVGF 666
cotton vicilin      588 ASE                   590
cocoa vicilin       515 VIKFTVKASAY           525
```

Fig. 6 (6/6)

```
                  1         10        20        30        40        47
                  |         |         |         |         |         |
MiAMP2c           RQRDPQQQYE QCQKRCQRRE TEPRHMQICQ QRCERRYEKE KRKQQKR

Gibrat method     CCCCCCCCCH HHECCCCCCC CCCCCCEEEC CCCCCCCHHH HHHHHHH
Levin method      CCCCCHCCHH HHHHHHCHHT HCSCCCCECC CHHHTHHHHH HHHHCHH
DPM method        CCCCCCCCCH HHHHHHHHHH CHCCCHHEEH HHHHHHHHHH HHHHHCC
SOPMA method      CCCCCHHHHH HHHHHEECCC CCCCHEEEEE EHHHHHHHHH HHHHHHH
PhD method        CCCCHHHHHH HHHHHHHHHH CCCCCHHHHH HHHHHHHHHH HHHHCCC

Consensus         CCCCCHCCHH HHHHHH-HH- CCCCC--EE- -HHHHHHHHH HHHHHHH
```

Fig. 7

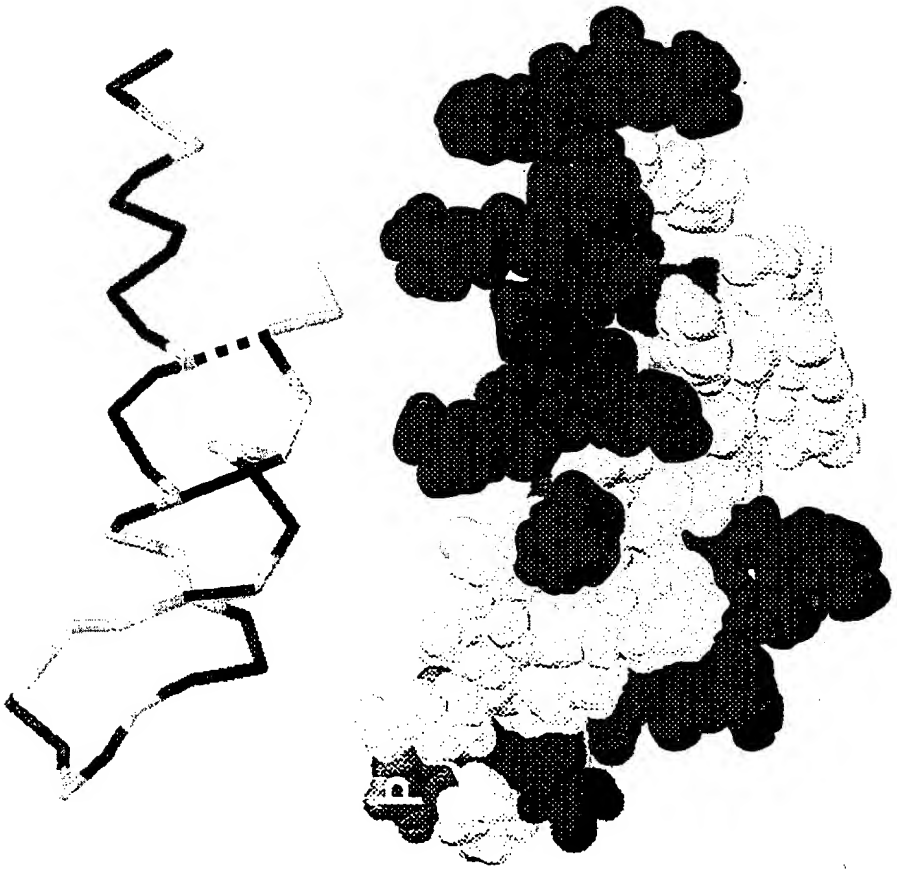

Fig. 8

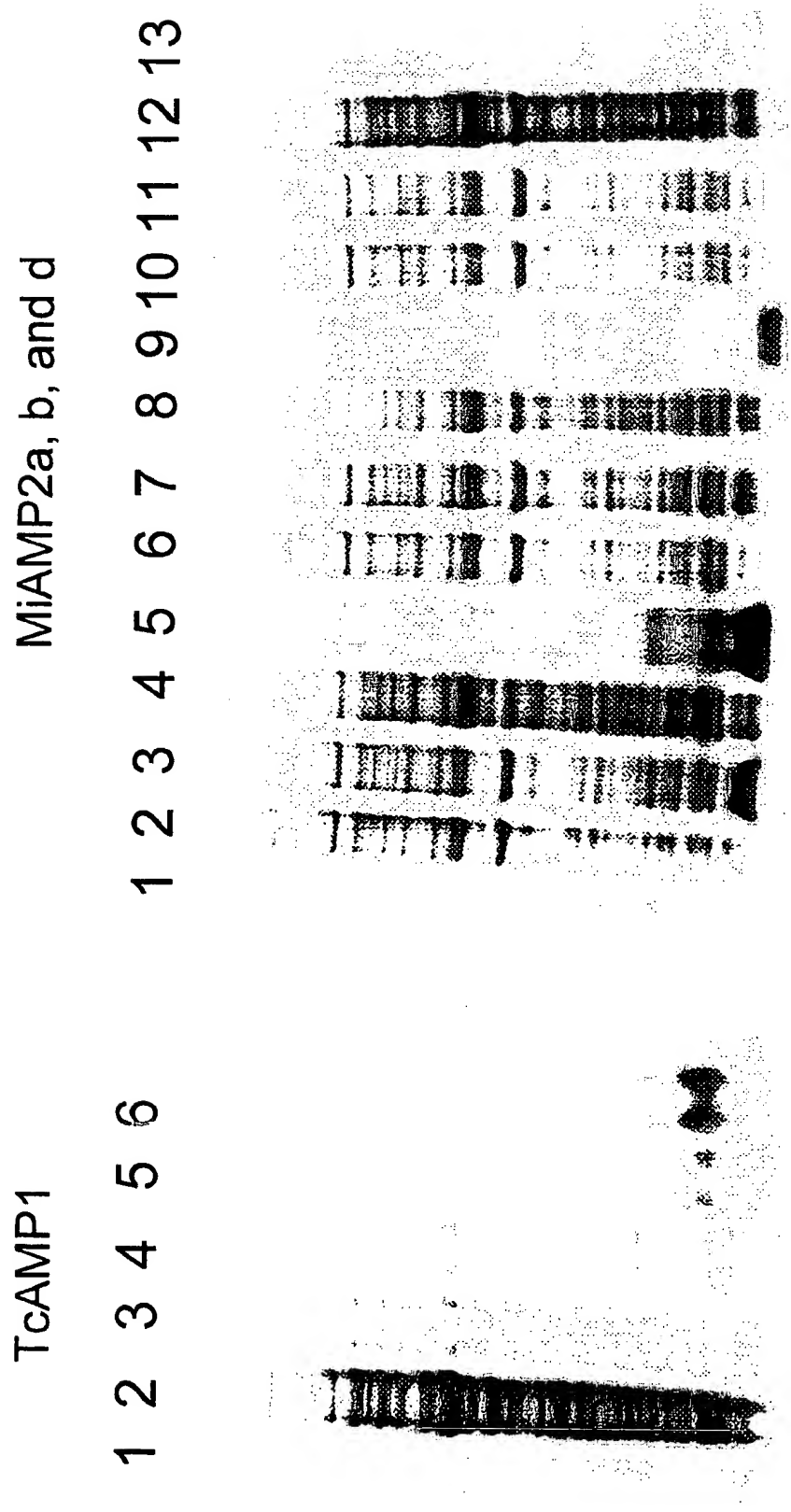

Fig. 9

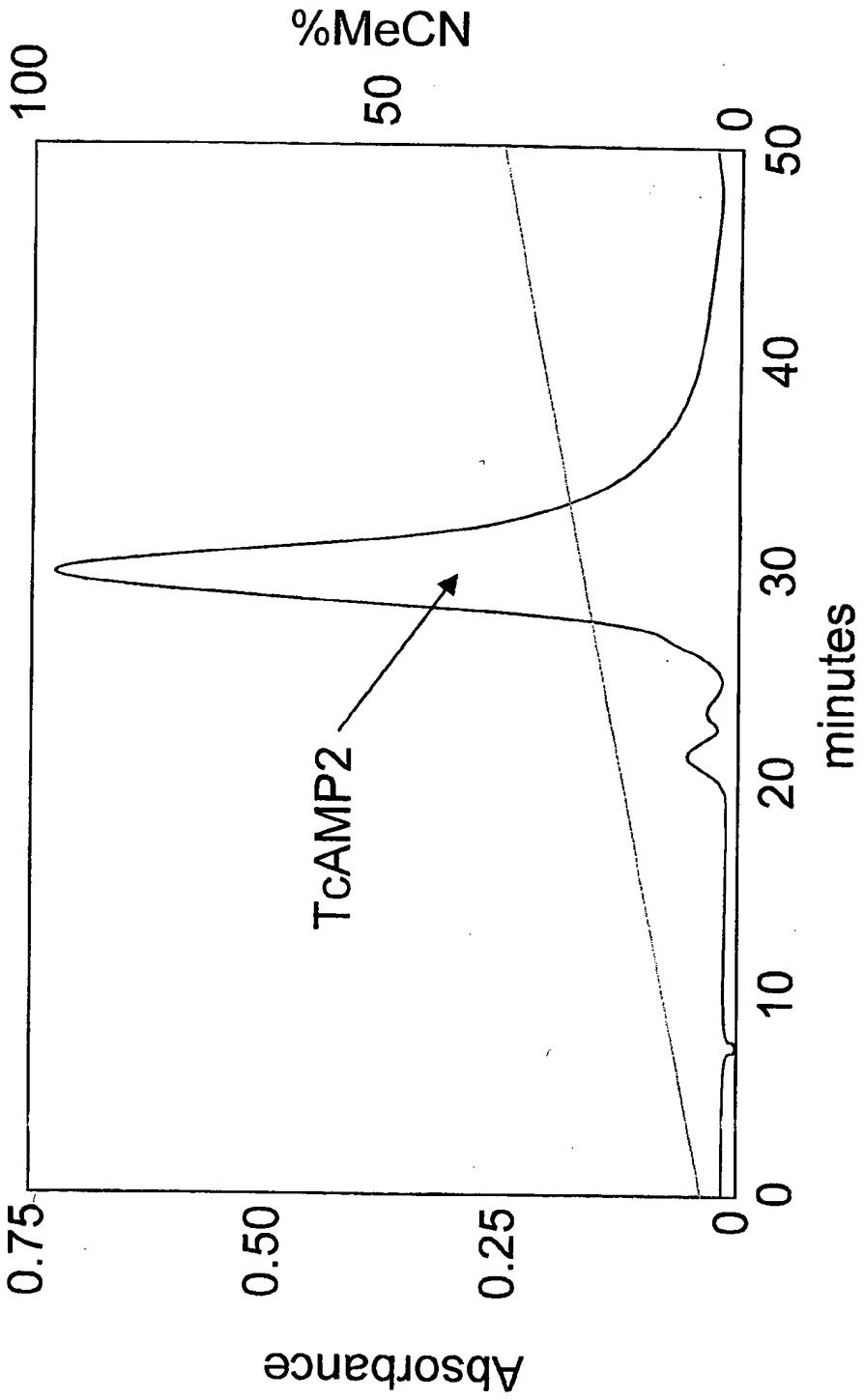
%MeCN
100
50
0
0
10
20
30
40
50
minutes
TcAMP2
0.75
0.50
0.25
0
Absorbance

Fig. 10

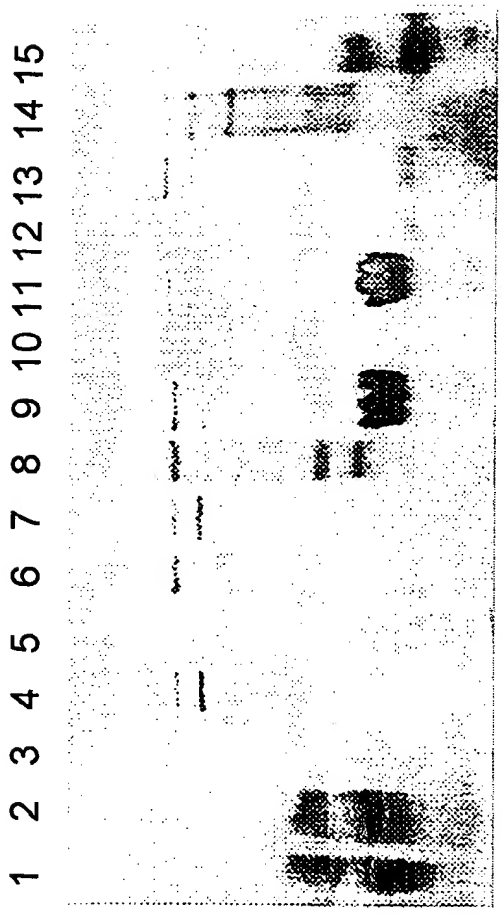

Fig. 11

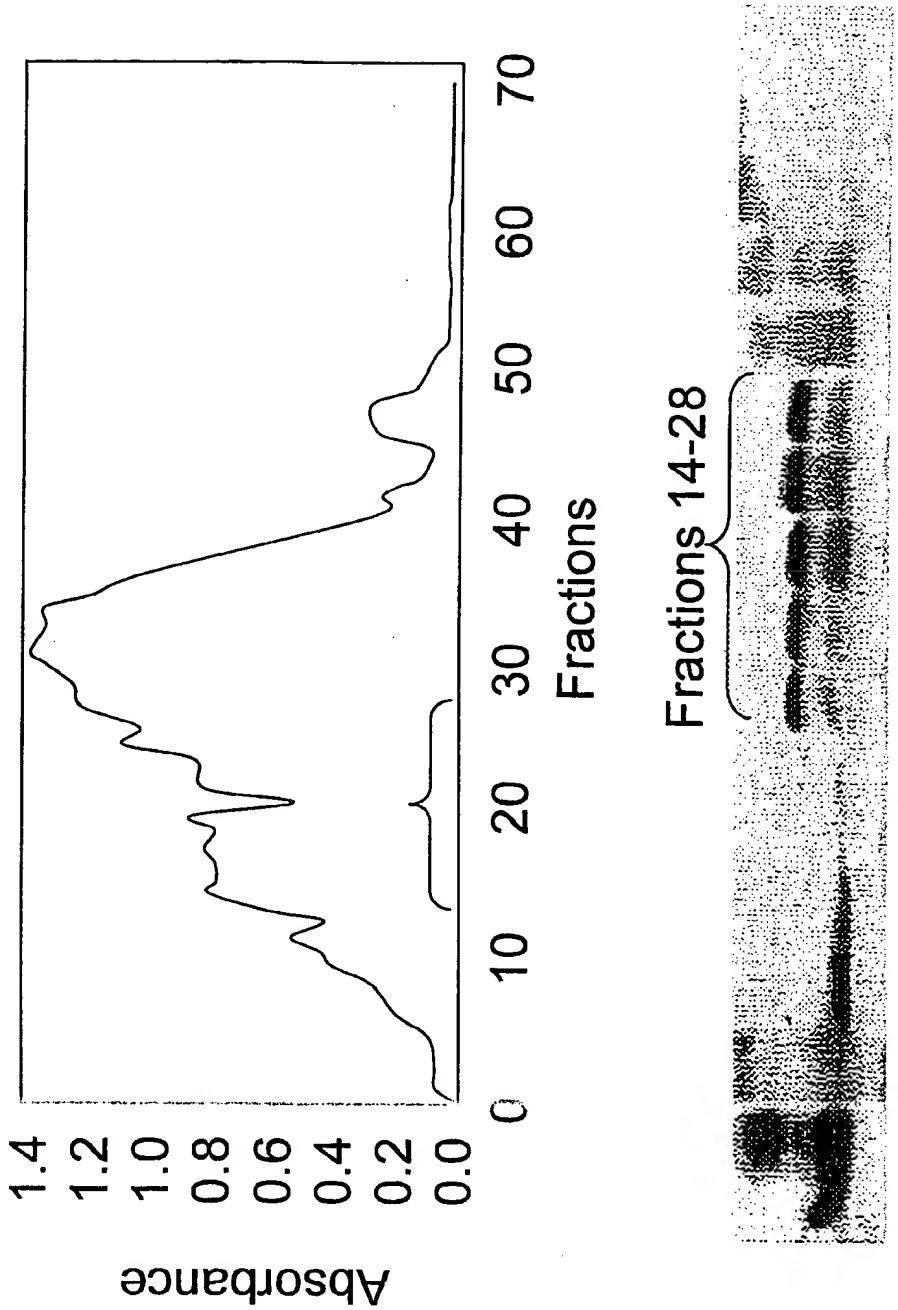
Absorbance
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
60
70
Fractions
Fractions 14-28

Fig. 12

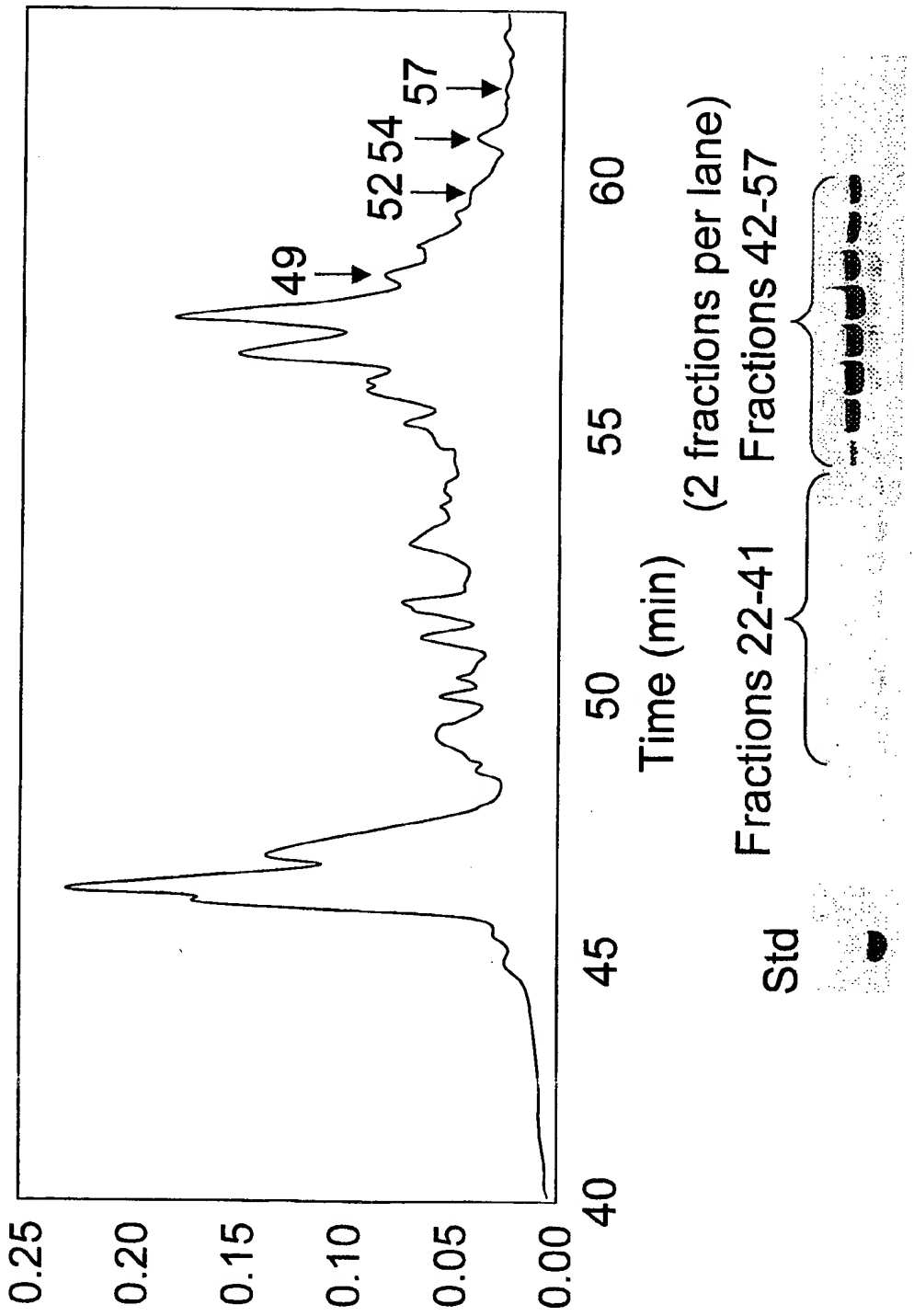
0.25
0.20
0.15
0.10
0.05
0.00
40
45
50
55
60
Time (min)
49
52
54
57
(2 fractions per lane)
Fractions 22-41
Fractions 42-57
Std

Fig. 13

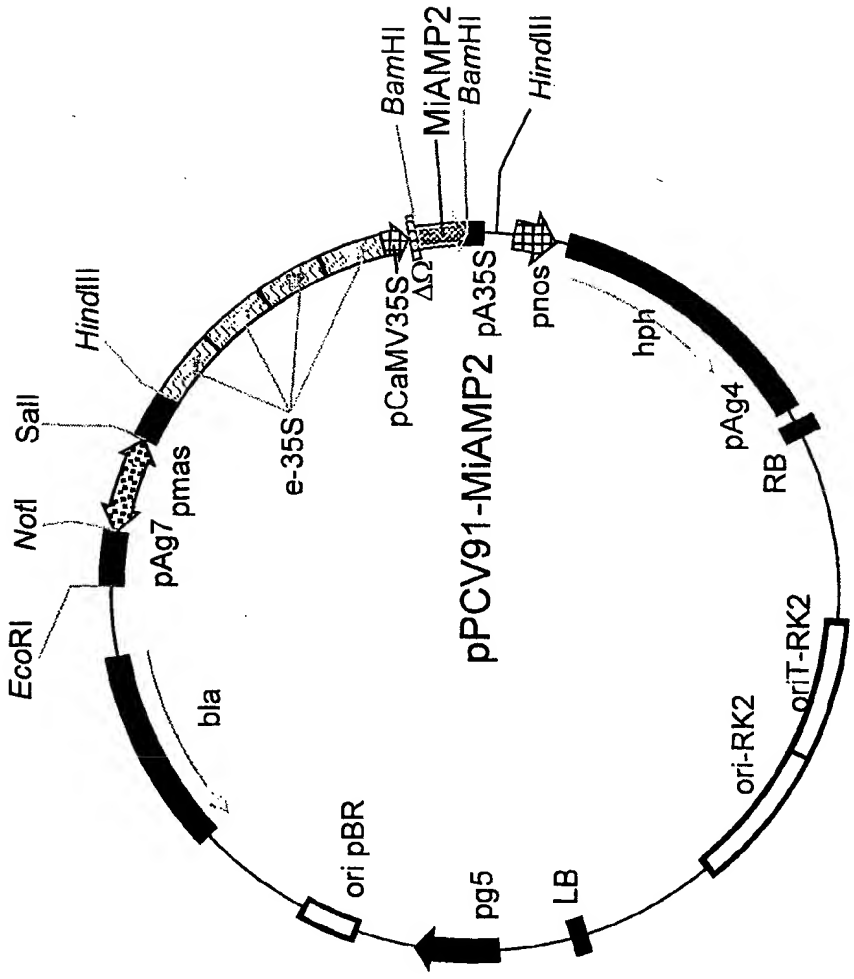
pPCV91-MiAMP2
BamHI
MiAMP2
BamHI
HindIII
pCaMV35S
ΔΩ
pA35S
pnos
hph
pAg4
RB
oriT-RK2
ori-RK2
LB
pg5
ori pBR
bla
EcoRI
NotI
SalI
pAg7
pmas
HindIII
e-35S

Fig. 14

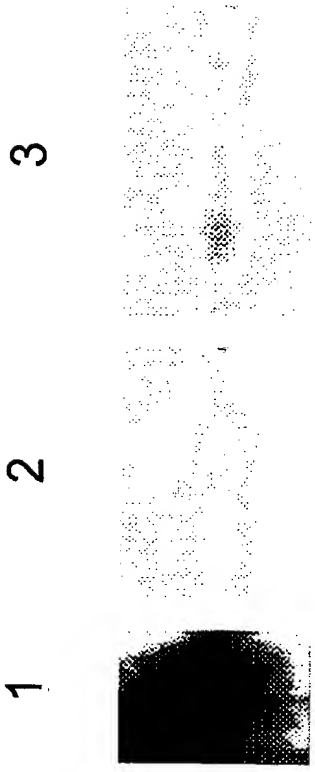

Fig. 15

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/AU 97/00874

A. CLASSIFICATION OF SUBJECT MATTER

Int Cl[6]: A01H 5/00, 5/10; A01N 37/18; C07K 4/10, 7/08, 14/415; C12N 15/29

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
CHEMICAL ABSTRACTS: See below

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
DGENE (Keywords as below)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
STN (Cas Online, DGENE) - CXXXCX(10, 12)CXXXC
Swiss Prot, Genbank, EMBL, PIR-SEQ IDs 1, 3, 5, 7, 8, 21, 22, 24, 25, 26, 27, 28

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 91/19801 (MARS UK LIMITED) 26 December 1991.	1, 4, 6, 11, 12
X	Plant Mol. Biol., Vol. 9, No. 6, 1987, Chlan et al., "Developmental biochemistry of cottonseed embryogenesis and germination XIX. Sequences and genomic organisation of the α-globulin (vicilin) genes of cottonseed", pages 533-46.	1, 4, 6, 11, 12
X	Plant Mol. Biol., Vol. 18, No. 6, 1992, McHenry and Fritz, "Comparison of the structure and nucleotide sequences of vicilin genes of cocoa and cotton raise questions about vicilin evolution", pages 1173-6.	1, 4, 6, 11, 12

[X] Further documents are listed in the continuation of Box C

[X] See patent family annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 February 1998	3 MAR 1998
Name and mailing address of the ISA/AU AUSTRALIAN INDUSTRIAL PROPERTY ORGANISATION PO BOX 200 WODEN ACT 2606 AUSTRALIA Facsimile No.: (02) 6285 3929	Authorized officer **CHRISTOPHER LUTON** Telephone No.: (02) 6283 2256

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/AU 97/00874

C (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Genetics, Vol. 129, No. 3, 1991, Belanger et al., "Molecular basis for allelic polymorphism of the maize Globulin-1 gene", pages 863-872.	1, 4, 6, 11, 12
X	Biochem. Genet., Vol. 27, No. 3-4, 1989, Kriz, "Characterization of embryo globulins encoded by the maize Glb genes", pages 239-251.	1, 4, 6, 11, 12
X	Mol. Gen. Genet., Vol. 239, No. 1-2, 1993, Heck et al., "Barley embryo globulin-1 gene, Beg1: characterization of cDNA, chromosome mapping and regulation of expression", pages 209-216.	1, 4, 6, 11, 12
X	J. Clin. Invest., Vol. 96, No. 4, 1995, Burks et al., "Recombinant peanut allergen Ara h1 expression and IgE binding in patients with peanut hypersensitivity", pages 1715-1721.	1, 4, 6, 11, 12
X	Plant Mol. Biol., Vol. 15, No. 1, 1990, Sebastiani et al., "Complete sequence of a cDNA of alpha subunit of soybean beta-conglycinin", pages 197-201.	1, 4, 6, 11, 12
X	Plant Mol. Biol., Vol. 7, No. 6, 1986, Chlan et al., "Developmental biochemistry of cottonseed embryogenesis and germination. XVIII. cDNA and amino acid sequences of members of the storage protein families.", pages 475-489.	1, 4, 6, 11, 12
X	TREMBL database entry, Accession No: Q41750, 1 November 1996.	1, 4, 6, 11, 12
X	EMBL database entry, Accession No: U28017, 7 August 1995.	1, 4, 6, 11, 12

INTERNATIONAL SEARCH REPORT

...ternational Application No.
PCT/AU 97/00874

Box 1 Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 1 (parts (vi) and (vii)) and 16
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
Due to the broad and indefinite scope of these claims, the International Search Authority finds that for economic reasons no meaningful search could be carried on said claims.

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International Application No.
PCT/AU 97/00874

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report		Patent Family Member					
WO	91/19801	AU,A,	79782/91	EP,A,	535053	GB,A,	9013016
		JP	5507846				

END OF ANNEX